

Db 8123 ATCATCGGACATTTAAAGACAGCGGTGCTCAAGCCTGTAATCCAGCACATTGGGAG 8064
Qy 1787 nnn 1846
Db 8063 GCTGAGCGGGAGGATCACCAGGTCAGGAGATCCAGACCATTCTGGCTACACGGTGAAA 8004
Qy 1847 nnn 1906
Db 8003 CCCCGTCTCTACTAAAAATACAAAAAATAGCTGGGCATGTTGGCGGCA---CCAGTCC 7947
Qy 1907 cagctactgggagcgtgagcgtgagcgtgagcgtgagcgtgagcgtgagcgtgagcgt 1966
Db 7946 AGGCTACTCAGGAGGCTGAGCGAGGAAATGGCGTGAACCCAGGAGCGGAGCTTGCAGT 7887
Qy 1967 gagcggagcggcgccactcactcagcgtgagcgtgagcgtgagcgtgagcgtgagcgt 2022
Db 7886 GAGCTGAGATGGCAGCTGCTGCATCCAGCCTGGGCAACAGAGTGAGACTCCATCTCAAT 7827
Qy 2023 ---aaaaaaagtaacttagtgagcgtgagcgtgagcgtgagcgtgagcgtgagcgt 2079
Db 7826 AATAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 7767
Qy 2080 agaccgtgccccgttatgaggtgtacagaaagcaagatttcaactatgcacactatc 2139
Db 7766 ACTTTGGGAGGCGGAGGCAAGCAGATCACCTAAGGCCAAGAGTTCAAGACCGCTGACC 7707
Qy 2140 accgctcaccttagcattgaagcagcgtgagcgtgagcgtgagcgtgagcgtgagcgt 2199
Db 7706 AACATGTGAACCCCACTCTACTAAAAATATTTTAAAAATATTTTAAAAATATTTTAAAA 7647
Qy 2200 tcttccccaaaatgcagacacgaagtgcaagtgcaagtgcaagtgcaagtgcaagtgcaag 2259
Db 7646 GCGCGCTCTAATCCCAGCTACTCAGGAGGCTGAGGAGGAGGAGGAGGAGGAGGAGGAG 7587
Qy 2260 taactgtcacgaagccagcagtgagcgtgagcgtgagcgtgagcgtgagcgtgagcgt 2319
Db 7586 GATGGAGGCTGCAGTGAGCGGAGATCACACATTTGCTCTCCAGCCTGGGTGACAGAGCCA 7527
Qy 2320 atgtgagcagcgggagcgggtcccttagagataagaga-----caataaagggga 2373
Db 7526 GACTCCCTCTCAACCAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAAC 7467
Qy 2374 gatacagagaaaatcgttaaggggagcagatggtgtcaagagaaataggctgaccatcga 2433
Db 7466 GTTCACAACGAATTGATCACAGCGAGTAGAATTTCTTCATTTCTTCCTCCAGTCTACTGC 7407
Qy 2434 aggaactggcgaagccttcgagaaacca---ctggagcgtgagcagtgagcgtgagcgt 2490
Db 7406 TTTGCTTTGACCGCTTTAAAGACACACATATATTTTGTCTGGCGGCTTGGCTCACACC 7347
Qy 2491 tgaatccccagcactttgagcgtgagcgtgagcgtgagcgtgagcgtgagcgtgagcgt 2550
Db 7346 TGTAAATCCCAACACTTTGGGAGGCCAAGGAGCGGATCAGCTGAGCTGAGGATTTGAG 7287
Qy 2551 accagcctggccacatggtgaaacccatctctacagaaaaataaaaaattagccagc 2610
Db 7286 ACCAGCCTGACCAACGTTGGAGAAACCCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 7228
Qy 2611 gtgggtgcacaaagcctagatccacagtaacttgggagcgtgagcgtgagcgtgagcgt 2654
Db 7227 ATGTTGGACATGCTGCTGTAATCCAGCTACTGGAGAGGCTGAGG 7184

RESULT 7
US-09-305-384-5
; Sequence 5, Application US/09305384
; Patent No. 6242218
; GENERAL INFORMATION:
; APPLICANT: Treco, Douglas A.
; APPLICANT: Heartlein, Michael W.
; APPLICANT: Selden, Richard F.
; TITLE OF INVENTION: GENOMIC SEQUENCES FOR PROTEIN PRODUCTION AND DELIVERY

; FILE REFERENCE: 07236/017001
; CURRENT APPLICATION NUMBER: US/09/305,384
; FILING DATE: 1999-05-05
; EARLIER APPLICATION NUMBER: US 60/084,649
; EARLIER FILING DATE: 1998-05-07
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 6235
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-305-384-5

Query Match 6.4%; Score 169.6; DB 4; Length 6235;
Best Local Similarity 54.4%; Pred. No. 5.1e-25;
Matches 442; Conservative 0; Mismatches 344; Indels 26; Gaps 4;

Qy 1869 aaaaattagctggcggtggcgatgcataccaaatccccagctactgggggagcgtgagcgt 1928
Db 207 aaaaattagccagcagcgtggcgagccacgttaatccggctactcagcagcgtgagcgt 266
Qy 1929 atgagaatcgttaaacgggggagcagatgttcagtgagcagcagcgtgagcgtgagcgt 1988
Db 267 aggaatcactgaacccagggcgaggtgagtgagtgagtgagtgagtgagtgagtgagtg 326
Qy 1989 ctcagcctggactacagagcagactctctcaaaaaaaataaaaaaaataaaaaaaataa 2048
Db 327 ctcagcctgggtgacagagcagactctctcaaaaaaaataaaaaaaataaaaaaaataa 386
Qy 2049 ttgggtgcaggt-----gtcctctgtttactcagcgtgagcgtgagcgtgagcgtgagcgt 2102
Db 387 agcagcagcgtgtagtcacaccccttagctcagctactcagcgtgagcgtgagcgtgagcgt 446
Qy 2103 tgtaccagaaagcaagtattc-actatgcacactattcaccgctcacccctcagcgtgagcgt 2161
Db 447 tcacttgaacctgggcagtcgaagtcacagtcagtcagtcagtcagtcagtcagtcagtc 506
Qy 2162 ccagcctgtagcctgaaagccttggctggggcaggtcttccccaaaaatgcagacac 2221
Db 507 cctgggcaacagagagacccctgtctcaaaaaaaataaaaaaaataaaaaaaataaaaaa 566
Qy 2222 gaaggtgcaagtgaaagcgtgagcgtgagcgtgagcgtgagcgtgagcgtgagcgtgagcgt 2281
Db 567 cctgtttatgtctcctggtccacatactactatgtatatagtttgcacaaactcaaaaga 626
Qy 2282 gtgcagggagagcgtgtcccaatttcggaagt---ggctatgtgagcagcgtgagcgtgagcgt 2338
Db 627 tccagatagtcatttttttaggtgtggcgctgagtgctctgtcacaatcactctgccc 686
Qy 2339 ggtcccttagagataagagacaaatcaaggggagata-----tcaga 2382
Db 687 ctgcttcttagcaaaaagcagcgtataaacaatacacaatacagaattttttatagacatc 746
Qy 2383 gaaaatcgttaaggggagcagatggtgtcagaagaaataggctgaccatcgaagcgtgagcgt 2442
Db 747 gagatttgaatttcataatgtttttacattttataaaataatcttttaaaaaattttccc 806
Qy 2443 agaagccttcagaaaaaccactggcggtggcgagcgtgagcgtgagcgtgagcgtgagcgt 2502
Db 807 ctacccatttaaaagttaaaagcggcgagcgtgagcgtgagcgtgagcgtgagcgtgagcgt 866
Qy 2503 acttgggagcgtgagcgtgagcgtgagcgtgagcgtgagcgtgagcgtgagcgtgagcgtgagcgt 2562
Db 867 acttgggagcgtgagcgtgagcgtgagcgtgagcgtgagcgtgagcgtgagcgtgagcgtgagcgt 926
Qy 2563 aacatggtgaacccccatctctacagaaaaataaaaaattagccagcgtgagcgtgagcgtgagcgt 2622
Db 927 aacatgcaaaaacccatttctactaaaaataaaaaaaattagcgtgagcgtgagcgtgagcgtgagcgt 986
Qy 2623 gcttagaattccagcacttctgggagcgtgagcgtgagcgtgagcgtgagcgtgagcgtgagcgt 2654
Db 987 accctgtatccagcacttctgggagcgtgagcgtgagcgtgagcgtgagcgtgagcgtgagcgt 1018

QY 2503 actttggaggctgacgagtgtaactcacttgagtgaggttcacagaccagcctggcc 2562
|||||
Db 886 actttggaggctgagtgaggcagatcacttgagatcaacagttcagaccagcctggcc 945
QY 2563 aacatgggtgaaccccatctctacagaaaataaaaaattagccagcggtggtggcaca 2622
|||||
Db 946 aacatgacaaaccccatctctactcaaaaaataaaattagctgggcatagtggtgcac 1005
QY 2623 gctagaatcccgactctacttgaggctgagg 2654
|||||
Db 1006 acctgtgattcccgactctacttgaggctgagg 1037

RESULT 9
US-08-480-784-20
; Sequence 20, Application US/08480784
; Patent No. 5693473
; GENERAL INFORMATION:
; APPLICANT: Skolnick, Mark H.
; APPLICANT: Goldgar, David E.
; APPLICANT: Miki, Yoshio
; APPLICANT: Swenson, Jeff
; APPLICANT: Kamb, Alexander
; APPLICANT: Harshman, Keith D.
; APPLICANT: Shattuck-Eidens, Donna M.
; APPLICANT: Tavtigian, Sean V.
; APPLICANT: Wiseman, Roger W.
; APPLICANT: Futreal, P. Andrew
; TITLE OF INVENTION: 17q-Linked Breast and Ovarian Cancer
; NUMBER OF SEQUENCES: 85
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Venable, Baetjer, Howard & Civiletti, LLP
; STREET: 1201 New York Avenue, N.W., Suite 1000
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/480,784
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/409,305
; FILING DATE: 24-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/348,824
; FILING DATE: 29-NOV-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/308,104
; FILING DATE: 16-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/300,266
; FILING DATE: 02-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/289,221
; FILING DATE: 12-AUG-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Ihnen, Jeffrey L.
; REGISTRATION NUMBER: 28,957
; REFERENCE/DOCKET NUMBER: 24884-109347
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-962-4810
; TELEFAX: 202-962-8300
; INFORMATION FOR SEQ ID NO:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6769 base pairs
; TYPE: nucleic acid

US-09-305-384-1
; Sequence 1, Application US/09305384
; Patent No. 6242218
; GENERAL INFORMATION:
; APPLICANT: Treco, Douglas A.
; APPLICANT: Heartlein, Michael W.
; APPLICANT: Selden, Richard F.
; TITLE OF INVENTION: GENOMIC SEQUENCES FOR PROTEIN PRODUCTION AND DELIVERY
; FILE REFERENCE: 07236/017001
; CURRENT APPLICATION NUMBER: US/09/305,384
; EARLIER FILING DATE: 1999-05-05
; EARLIER FILING DATE: 1998-05-07
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 6679
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-305-384-1

Query Match 6.4%; Score 169.6; DB 4; Length 6679;
Best Local Similarity 54.4%; Pred. No. 5.1e-25;
Matches 442; Conservative 0; Mismatches 344; Indels 26; Gaps 4;

QY 1869 aaaaattagctggcggtggtgcatgcacaaatcccgactctggtggagcgtgagcc 1928
|||||
Db 226 aaaaattagcagcgtggtggtgagccctgtaatcccgactctggtgagcgtgagcc 285
QY 1929 atggaatcgctgaacggggagggagacatgttgagtgagccgagagcgccactgca 1988
|||||
Db 286 aggaatcactgaacccagggagggaggttgagtgagcagatcagaccactgca 345
QY 1989 ctccagcctgactcagagcagactctctcaaaaaaataaaaaaagaagtaac 2048
|||||
Db 346 ctccagcctggtgacagagcagactctctcaaaaaaataaaaaaagaagtaac 405
QY 2049 ttagtgagcgggt-----gtcctctgttatcactgagacgctgcccgttatgaggt 2102
|||||
Db 406 agccagcgtggtgagcagcctctcagactctcagactcagagcgtgaggtggagga 465
QY 2103 tgtaccagaaagcaagtatc-actatgcacactattcaccgctcaccctagcattgaag 2161
|||||
Db 466 tcaactgaacctggggcagtgcaaggctacagtgcagcaagatcagccactacactccag 525
QY 2162 caagcctgtagcctgaaagccttctgttgaggcaggtcttcccaaaaatgcagacac 2221
|||||
Db 526 cctgggcaacagagagagacccctgtctcaaaaaaataataataaagaaaaaacag 585
QY 2222 gaagtgcaagtgaaagctgcagctcttgcaaaagatgtaactgtgtcacgaagccacga 2281
|||||
Db 586 ctctgtttatgtctcctggctccatcacataactactatgtatagtttgcaaaactcaaga 645
QY 2282 gtggcaggagagcgtgtccacatttgcgaagt---ggctatgtgagcagggggagggc 2338
|||||
Db 646 tccagatagtaatttttagctgtg1999ccgtgtgctctgtcacaatcactctgccc 705
QY 2339 ggggtcccttagatagagacaatcataaggggagata-----tcaga 2382
|||||
Db 706 ctgtcttcttagcacaagaagcagctataaacaataacatacatgaattttttatagaatc 765
QY 2383 gaaatcgtaaggaggagcagatggttgcagaagaataggtcgaccatcgaaagcactggc 2442
|||||
Db 766 gagatttgaaatttcataatgattttttacattttataaataaattcttttaaaattttccc 825
QY 2443 agaagcttccagaaacacactgcagcgtggtggcagcagtggtgagcagtgctgtaattccagc 2502
|||||
Db 826 ctaaccatttaaaagttaaaagccggccagcgccatcgccagcgtcagcgtcgaattccagc 885

```

: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: HYPOTHETICAL: NO
: ANTI-SENSE: NO
: ORIGINAL SOURCE:
: ORGANISM: Homo sapiens
: US-08-480-784-20

Query Match          6.2%; Score 165.4; DB 1; Length 6769;
Best Local Similarity 55.8%; Pred. No. 3.4e-24;
Matches 464; Conservative 0; Mismatches 321; Indels 46; Gaps 6;

Qy 1869 aaaaattagctggcggtggatgcatacccaactccagctactgggaggtgagc 1928
Db 160 AAAAAATTAGTGGTGGTGGCAGCTACCTGTAAATCCAGCTACTCGGAGGCGGAGGC 219
Qy 1929 atgagaatcgttgaacccgggagcgagatgttgcagtgagccgagcgcgcactgca 1988
Db 220 ACAAGAATTGCTTGAAGCTAGGAGCGGAGGTTGCAGCGAGCAAGATCGCGCCACTGCA 279
Qy 1989 ctccagctgagactacagcagacactctctcaaaaaaataaaaaaagaat-a 2046
Db 280 CTCCAGCCTGGCGCGTAGAGTGAGACTCTGTCTCAAAAAAGAAAAAAGTAATTGTCT 339
Qy 2047 acttagtgaggggtgctctctgttattcactgagaccgtgcccgggttagagtt--- 2103
Db 340 AGCTGGGCGCAGTGGCTTTCCTCTGTAAATCCACACACTTTGGGAGGCCAAGCGCGGTGA 399
Qy 2104 -----gtaccagaaagaagtattcactatgcacattatccacgtcacccttagcatt 2157
Db 400 TCTCGAGTCTCTAGAGTTCAAGACCAGCCTAGGCAATGTGTGAACCCCACTCGCTACAAA 459
Qy 2158 gaagcagcctgtagctgaagccttctgtttagagcaggtttttccc-----c 2208
Db 460 AAATACAAAAAATTAGCCAGGCGATGGTGGCGTCCGATGTAGTCCAGCTCTTGGGAGGC 519
Qy 2209 aaaaatgcagacacgaagtgcaagtgcaagtgcaagtgcaagtgcaagtgcaagtgca 2268
Db 520 TGAGGTGGGAGGATCACTTGAACCCAGGAGACAGAGTTGCAGTGAACCGAGATCAGGCC 579
Qy 2269 acgaagggccagagtgccagggagagctgttccacatttgcggagtgctat----- 2321
Db 580 ACCACGCTCCAGCCTGGGCAACAGCAAGACTCTGTCTAAAAAATAACAAATAAATAA 639
Qy 2322 -----gtgagacggggagggcggtcccttagagataagagacaatacgaaggagat 2376
Db 640 AAGTAGTCTCACAGTACCAGCATTCATTTTCAAAAGATATAGAGCTAAAAAGGAAGGA 699
Qy 2377 atcagagaaaatcgttaaggggagcagatggtgtcagagagataagctgaccatcgaagg 2436
Db 700 AAAAAAAGTAATCTGGGCTTTTAAATACTCTGTTCTTAAATCTTCTTAGGAGTG 759
Qy 2437 actggcagaagcttcca-----gaaacacactggagcggctgggacagtgagc 2483
Db 760 CTGGGCTTTTATTTGCTCATTTATCTCTTTTAAAAATGTTATTGGCAGGACGCGTGGC 819
Qy 2484 ttaggctgtaaatccagcacttggagcgtgacgaggtgaatcacttgaagtcagga 2543
Db 820 TCATGGCTGTANTCCAGCACTTTGGAGGCGGAGGAGGAGATCCTCAGGCTCAGGA 879
Qy 2544 gtccagaccagcctggccacaatggtgaaaccccatctctacagaaaaataaaaaatta 2603
Db 880 GTGTGAGACCAAGCTGGCCACATGGCGAAACCTG-TCTCTACTAAAAATAAATAAATTA 938
Qy 2604 gccagggtggtgacacaagcctagaatccccagctacttgggaggtcaggg 2654
Db 939 ACTAGGCGTGTGTACGCGCTGTAGTCCAGCTACTCGCGGAGGCTCAGG 989

RESULT 10
US-08-483-553-20
```

```

: Sequence 20, Application US/08483553
: Patent No. 5709999
: GENERAL INFORMATION:
: APPLICANT: Skolnick, Mark H.
: APPLICANT: Goldgar, David E.
: APPLICANT: Miki, Yoshio
: APPLICANT: Swenson, Jeff
: APPLICANT: Kamb, Alexander
: APPLICANT: Harshman, Keith D.
: APPLICANT: Shattuck-Eidens, Donna M.
: APPLICANT: Tavtigian, Sean V.
: APPLICANT: Wiseman, Roger W.
: APPLICANT: Futreal, P. Andrew
: TITLE OF INVENTION: 17q-Linked Breast and Ovarian Cancer
: NUMBER OF SEQUENCES: 85
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Venable, Baetjer, Howard & Civiletti, LLP
: STREET: 1201 New York Avenue, N.W., Suite 1000
: CITY: Washington
: STATE: DC
: COUNTRY: USA
: ZIP: 20005
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/483,553
: FILING DATE:
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/409,305
: FILING DATE: 24-MAR-1995
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/348,824
: FILING DATE: 29-NOV-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/308,104
: FILING DATE: 16-SEP-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/289,221
: FILING DATE: 12-AUG-1994
: ATTORNEY/AGENT INFORMATION:
: NAME: Ihnen, Jeffrey L.
: REGISTRATION NUMBER: 28,957
: REFERENCE/DOCKET NUMBER: 24884-109347
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 202-962-4810
: TELEFAX: 202-962-8300
: INFORMATION FOR SEQ ID NO: 20:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 6769 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: HYPOTHETICAL: NO
: ANTI-SENSE: NO
: ORIGINAL SOURCE:
: ORGANISM: Homo sapiens
: US-08-483-553-20
```

```

Query Match          6.2%; Score 165.4; DB 1; Length 6769;
Best Local Similarity 55.8%; Pred. No. 3.4e-24;
Matches 464; Conservative 0; Mismatches 321; Indels 46; Gaps 6;

Qy 1869 aaaaattagctggcggtggatgcatacccaactccagctactgggaggtgagc 1928
```

```
Db 160 AAAAAATTAGCTGGGTGTGGTGGTACCTGTATATCCAGCTACTCGGGAGCGGAGGC 219
QY 1929 atgagatcgttgaaacgggagagcagatgttgcagtgcagcagcagcagcagcagc 1988
Db 220 ACAAGAAATTCCTTGAACCTTAGGACGCGGAGGTGGAGCGCAAGATCGCCACTGCA 279
QY 1989 ctccagcctggactacagacgagacgtctctcaaaaaaaataaaaaaaagt--a 2046
Db 280 CTCCAGCCTGGCCGTAGAGTGAAGTCTGTCTCAAAAAGAAAGAAAGTAAATGTCT 339
QY 2047 acttagtgaggggtgtcctctgttattcaatgcagcagcagcagcagcagcagc 2103
Db 340 AGCTGGCGCAGTGGCTCTGTGCTGTATATCCAGCACTTTGGGAGGCCAAGCGGTGGA 399
QY 2104 -----gtaccagaagaagattcaatgcagcagcagcagcagcagcagcagc 2157
Db 400 TCTCGAGTCTAGAGTTCAGACGCTAGGCAATGTGGTGAACCCCACTCGCTACAAA 459
QY 2158 gaagccagcctgtagcctgaagccttctgttgcagcagcagcagcagcagcagc 2208
Db 460 AAATACAAAATAGCCAGCAGTGTGGCGTGCATGTAGTCCAGCTCTCTGGGAGGC 519
QY 2209 aaaaatgcagacacgaagtgcaaatgaagtcgagtcgagtcgagtcgagtcgagtc 2268
Db 520 TGAGGTGGGAGGATCACTTTGAACCCAGGAGACAGAGTTCAGTGAACCCGAGATCACGCC 579
QY 2269 acgaagggccagcagtgagcagcagcagcagcagcagcagcagcagcagcagcagc 2321
Db 580 ACACGCTCCAGCCTGGGCAACAGAACAGACTCTGTCTAAAAAATACAAAATAAAATA 639
QY 2322 -----gtgagggcggggagggcggccttagagataagagacaatcataagggagat 2376
Db 640 AAGTAGTCTACAGTACCAGCAATTCATTTTCAAAAGATATAGAGCTAAAAGGAAGA 699
QY 2377 atcagagaaaatcgtaagggagcagatggttgcagagagagagagagagagagag 2436
Db 700 AAAAAAAGTAAATGTGGGCTTTAAATACCTGTTCTTAAATGTTCTTAGGAGCTG 759
QY 2437 actgagcagaagcttca-----gaaacacactggagcggcgtggcagcagtgcc 2483
Db 760 CTGGGGTTTATGTCATCATTTATCTTTTAAATGTTTATGGCCAGCAGCGGTGGC 819
QY 2484 ttaggcctgaatccagcagccttggggagcgtgcagcagcagcagcagcagcagcagc 2543
Db 820 TCATGGCTGTAATCCAGCACTTTGGGAGCGCGGAGGAGGAGGAGGAGGAGGAGGAGG 879
QY 2544 gttcagacacgctggcgaacatggtgaacacccatctctcagagaaaataaaaaatta 2603
Db 880 GTGTGAGACCAAGCCTGGCCACATGGCGAAACCTG-TCTCTACTAAAAATACAAAAATTA 938
QY 2604 gccagggcgtggtggcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2654
Db 939 ACTAGGCGTGTGTGTAGCCTGTAGTCCAGCTACTCGGGAGCGGTGAGG 989
```

RESULT 11

US-08-487-002-20

Sequence 20, Application US/08487002

Patent No. 5710001

GENERAL INFORMATION:

APPLICANT: Shattuck-Eidens, Donna M.

APPLICANT: Simard, Jacques

APPLICANT: Emi, Mitsuru

APPLICANT: Nakamura, Yusuke

APPLICANT: Durocher, Francine

TITLE OF INVENTION: 17q-Linked Breast and Ovarian Cancer

TITLE OF INVENTION: Susceptibility Gene

NUMBER OF SEQUENCES: 85

CORRESPONDENCE ADDRESS:

ADDRESSEE: Venable, Baetjer, Howard & Civiletti, LLP

STREET: 1201 New York Avenue, N.W., Suite 1000

CITY: Washington

```
STATE: DC
COUNTRY: USA
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/487,002
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/409,305
FILING DATE: 24-MAR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/348,824
FILING DATE: 29-NOV-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/308,104
FILING DATE: 16-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/300,266
FILING DATE: 02-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/289,221
FILING DATE: 12-AUG-1994
ATTORNEY/AGENT INFORMATION:
NAME: Ihnen, Jeffrey L.
REGISTRATION NUMBER: 28,957
REFERENCE/DOCKET NUMBER: 24884-109347
TELEPHONE: 202-962-4810
TELEFAX: 202-962-8300
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 6769 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
US-08-487-002-20
```

Query Match 6.2%; Score 165.4; DB 1; Length 6769;

Best Local Similarity 55.8%; Pred. No. 3.4e-24;

Matches 464; Conservative 0; Mismatches 321; Indels 46; Gaps 6;

QY 1869 aaaaattagctggcgtgtgtggcagtcacacatcccgactactgggagggctgagc 1928

Db 160 AAAAAATTAGCTGGGTGTGGTGGCAGCTACCTGTATATCCAGCTACTCGGGAGCGGAGGC 219

QY 1929 atgagaatccttgaaccggggaggagcagatgttgcagtgcagcagcagcagcagc 1988

Db 220 ACAAGAAATTCCTTGAACCTTAGGACGCGGAGGTTCAGCAGCAAGATCGCCACTGCA 279

QY 1989 ctccagcctggactacagacgagacgtctctcaaaaaaaataaaaaaaagt--a 2046

Db 280 CTCAGCCTGGCCGTAGAGTGAAGTCTGTCTCAAAAAGAAAGAAAGTAAATGTCT 339

QY 2047 acttagtgaggggtgtcctctgttattcaatgcagcagcagcagcagcagcagc 2103

Db 340 AGCTGGCGCAGTGGCTCTTTCCTGTATATCCAGCACTTTGGGAGGCCAAGCGGTGGA 399

QY 2104 -----gtaccagaagaagattcaatgcagcagcagcagcagcagcagcagc 2157

Db 400 TCTCGAGTCTAGAGTTCAGACGCTAGGCAATGTGGTGAACCCCACTCGCTACAAA 459

QY 2158 gaagccagcctgtagcctgaagccttctgttgcagcagcagcagcagcagcagc 2208

```
Db 460 AAATACAAAAATTAGCCAGGCGATGGTGGCGTGCATGTAGTCCAGCTCCTTTGGGAGGC 519
Qy 2209 aaatgacagacgaaggtgcaaaagtgaagtgcaagctcttgcaaaagatgtaactgtgc 2268
Db 520 TGAGTGGGAGGATCACTTGAACCCAGGAGACAGAGGTGTGAGTGAACCGAGATCACGCC 579
Qy 2269 acgaagccacagtgagcgagggagagctgtcccacatttgcggagtggtctat----- 2321
Db 580 ACCAGCTCCAGCCTGGGCAACGAACAAGACTCTGTCTAAAAAATAACAAATAAATAA 639
Qy 2322 -----gtgagagcggggagggcgggtcccttagagataaagagcaataaagggagat 2376
Db 640 AAGTAGTCTCACAGTAGTCCAGCATCTTTTCAAAAAGATATAGAGCTAAAAAGGAAGA 699
Qy 2377 atcagagaaatcgtaaaggggagcagatggtgtcgaagagaataggctgacccatcgaag 2436
Db 700 AAAAAAAGTAAGTGGGCTTTTAAATACTCGTCTCTATCTAAATGTTCTTAGGAGTG 759
Qy 2437 actggcagaagctttca-----gaaacacactggagcgctgggcacagtggc 2483
Db 760 CTGGGCTTTTATTCATCATTTTATCTCTTTTAAAAATGTTATTGGCCAGGACGGTGGC 819
Qy 2484 ttaggctgtaatccccagcactttgggaggtgacgcaggtgaatcacttgaggtcagga 2543
Db 820 TCATGGCTGTAAATCCAGCACTTTGGGAGGCCGAGGAGGAGCATCACTGAGGTGACGA 879
Qy 2544 gtccagacagcctggtcccaactggtggaaccccatctctacagaaataataaaaaa 2603
Db 880 GTGTGACACAGCCTGGCCACATGCGGAACCTGT-TCTCTACTAAAAATAACAAAAATTA 938
Qy 2604 gccagcggtggtgcacaaagcctagaatccacagctacttgggagcgtgag 2654
Db 939 ACTAGGCTGTGTGTACGCTGTAGTCCAGCTACTTCGGGAGGCTGAGG 989
```

RESULT 12

US-08-483-554B-20
Sequence 20, Application US/08483554B
Patent No. 5747282

GENERAL INFORMATION:

APPLICANT: Skolnick, Mark H.
APPLICANT: Goldgar, David E.
APPLICANT: Miki, Yoshio
APPLICANT: Swenson, Jeff
APPLICANT: Kamb, Alexander
APPLICANT: Harshman, Keith D.
APPLICANT: Shattuck-Bidens, Donna M.
APPLICANT: Tavtigian, Sean V.
APPLICANT: Wiseman, Roger W.
APPLICANT: Futreal, P. Andrew
TITLE OF INVENTION: 17q-Linked Breast and Ovarian Cancer
TITLE OF INVENTION: Susceptibility Gene

NUMBER OF SEQUENCES: 85

CORRESPONDENCE ADDRESS:

ADDRESSEE: Venable, Baetjer, Howard & Civiletti, LLP

STREET: 1201 New York Avenue, N.W., Suite 1000

CITY: Washington

STATE: DC

COUNTRY: USA

ZIP: 20005

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/483,554B

FILING DATE: 07-JUN-1995

CLASSIFICATION: 514

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/409,305

FILING DATE: 24-MAR-1995

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/348,824
FILING DATE: 29-NOV-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/308,104
FILING DATE: 16-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/300,266
FILING DATE: 02-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/289,221
FILING DATE: 12-AUG-1994
ATTORNEY/AGENT INFORMATION:
NAME: Ihnen, Jeffrey L.
REGISTRATION NUMBER: 28,957
REFERENCE/DOCKET NUMBER: 24884-109347
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-962-4810
TELEFAX: 202-962-8300
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 6769 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
US-08-483-554B-20

Query Match 6.2%; Score 165.4; DB 1; Length 6769;

Best Local Similarity 55.8%; Pred. No. 3.4e-24;

Matches 464; Conservative 0; Mismatches 321; Indels 46; Gaps 6;

```
Qy 1869 aaaaattagctggcggtggtgcatccacaaatccacagctactggggaggtgaggc 1928
Db 160 AAAAAATTAGCTGGGTGGTGGCAGCTACCTGTAATCCAGCTACTCGGGAGGCGGAGGC 219
Qy 1929 atgagaatcgttgaacccggggagggagagatgttcagtgagccgagagcggccactgca 1988
Db 220 ACAAGAATTGCTTGAACCTAGGACGCGGAGGTTGAGGAGCAGCAATCGGCCACTGCA 279
Qy 1989 ctccagcctgactacagagcagactctatctcaaaaaaataaaaaaagtt--a 2046
Db 280 CTCAGCCTGGGCGGTAGAGTGAGACTCTGTCTCAAAAAAGAAAAAAGTAATTGTTCT 339
Qy 2047 acttagtgaggggtgtctctgttattcactgagaccgtgcccgggttatgaggt--- 2103
Db 340 AGCTGGGCGCAGTGGCTCTTCCCTGTAATCCAGCACATTTGGGAGGCCAAGCGGTTGGA 399
Qy 2104 -----gtaccagaaagcaagtattcactatgcacactattcaccgctcacccctagcatt 2157
Db 400 TCTCGAGTCTCTAGAGTTTCAAGACCAGCCTTAGGCAATGTGGTGAACCCCATCGCTACAA 459
Qy 2158 gaagccagcctgtagcctgaaagccttggttgagggcaggtcttctccc-----c 2208
Db 460 AAATACAAAAATTAGCCAGGATGGTGGCGGTAGTGTCCAGCTCTCTTGGGAGGC 519
Qy 2209 aaaaatgcagacacgaaggtgcaaaagtgaagctcttgcaaaagatgttaactgttc 2268
Db 520 TGAGGTGGGAGGATCACTTGAACCCAGGAGACAGAGGTGTCAGTCAACCGAGATCACGCC 579
Qy 2269 acgaagccacagtgagcgagggagagctgtcccacatttgcggaggtggtctat----- 2321
Db 580 ACCAGCTCCAGCCTGGGCAACAGAACAGACTCTGTCTAAAAAATAACAAATAAATAA 639
Qy 2322 -----gtgaggaacgggggagggcggtcccttagagataaagagacaataaagggagat 2376
Db 640 AAGTAGTCTTCACAGTACCAGCATTCATTTTCAAAAGATATAGAGCTAAAAAGGAAGA 699
```



```

Qy 2604 gccaggcgtggtggccacagccttagaatcccagctaacttggggaggcctgagg
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 939 ACTAGGCGTGGTGTTACGCCTGTAGTCCAGCTACTCTCGGGAGGCTGAGG

RESULT 14
US-08-850-727-20
: Sequence 20, Application US/08850727
: Patent No. 6162897
: GENERAL INFORMATION:
: APPLICANT: Skolnick, Mark H.
: APPLICANT: Goldgar, David E.
: APPLICANT: Miki, Yoshio
: APPLICANT: Swenson, Jeff
: APPLICANT: Kamb, Alexander
: APPLICANT: Harshman, Keith D.
: APPLICANT: Shattuck-Eidens, Donna M.
: APPLICANT: Tavtigian, Sean V.
: APPLICANT: Wiseman, Roger W.
: APPLICANT: Futreal, P. Andrew
: TITLE OF INVENTION: 17q-Linked Breast and Ovarian Cancer
: TITLE OF INVENTION: Susceptibility Gene
: NUMBER OF SEQUENCES: 85
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Venable, Baetjer, Howard & Civiletti, LLP
: STREET: 1201 New York Avenue, N.W., Suite 1000
: CITY: Washington
: STATE: DC
: COUNTRY: USA
: ZIP: 20005
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/850,727
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/483,554
: FILING DATE: 07-JUN-1995
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/348,824
: FILING DATE: 29-NOV-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/308,104
: FILING DATE: 16-SEP-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/300,266
: FILING DATE: 02-SEP-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/289,221
: FILING DATE: 12-AUG-1994
: ATTORNEY/AGENT INFORMATION:
: NAME: Ihnen, Jeffrey L.
: REGISTRATION NUMBER: 28,957
: REFERENCE/DOCKET NUMBER: 24884-109347
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 202-962-4810
: TELEFAX: 202-962-8300
: INFORMATION FOR SEQ ID NO: 20:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 6769 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: HYPOTHEetical: NO
: ANTI-SENSE: NO
: ORIGINAL SOURCE:
: ORGANISM: Homo sapiens

```

TITLE OF INVENTION: In Vivo Mutations and Polymorphisms
TITLE OF INVENTION: In the 17q-Linked Breast and Ovarian Cancer
NUMBER OF SEQUENCES: 85
CORRESPONDENCE ADDRESS:

ADDRESSEE: Venable, Baetjer, Howard & Civiletti, LLP
STREET: 1201 New York Avenue, N.W., Suite 1000
CITY: Washington
STATE: DC
COUNTRY: USA

ZIP: 20005

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US95/10202

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US

FILING DATE: 07-JUN-1995

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/409,305

FILING DATE: 24-MAR-1995

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/348,824

FILING DATE: 29-NOV-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08-308,104

FILING DATE: 16-SEP-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/300,266

FILING DATE: 02-SEP-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/289,221

FILING DATE: 12-AUG-1994

ATTORNEY/AGENT INFORMATION:

NAME: Ihnen, Jeffrey L.

REGISTRATION NUMBER: 28,957

REFERENCE/DOCKET NUMBER: 24884-109347

TELECOMMUNICATION INFORMATION:

TELEPHONE: 202-962-4810

TELEFAX: 202-962-8300

INFORMATION FOR SEQ ID NO: 20:

SEQUENCE CHARACTERISTICS:

LENGTH: 6769 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

HYPOTHETICAL: NO

ANTI-SENSE: NO

ORIGINAL SOURCE:

ORGANISM: Homo sapiens

PCT-US95-10202-20

Query Match 6.28; Score 165.4; DB 5; Length 6769;

Best Local Similarity 55.8%; Pred. No. 3.4e-24;

Matches 464; Conservative 0; Mismatches 321; Indels 46; Gaps 6;

QY 1869 aaaaattagctggcggtggtgcatgcatccacatcccaactactgagagcgtgagc 1928

Db 160 AAAAATTAGCTGGTGTGGTGGCAGCTACCTGTAACTCCACTACTCGGAGGCGGAGGC 219

QY 1929 atgagaatcgctgaacggggagcgagatgttgcagtgaagcgcgcgcactgca 1988

Db 220 ACAAGAATTGCTTCAACTAGGACGCGGAGGTTGCAGCGAGCCAAAGATCGCGCCACTGCA 279

QY 1989 ctcagccttgactacagagcgagactctatctcaaaaaaaaaaaaaaagt--a 2046

Db 280 CTCCAGCCTGGCCCTGAGAGTGAGACTCTGTCTCAAAAAAGAAAAAAGTAATTGTTCT 339

QY 2047 acttaggtgcaggtgtctctgttattcactgagacccgtgcccggttatgaggt--- 2103

Db 340 AGCTGGCGCAGTGGCTCTTGCTGTAAATCCAGCACATTTGGGAGGCCAAGCGGTTGGA 399

QY 2104 -----gtaccagaaagcaagtattcactatgacacattcaccgctcacccctagatt 2157

Db 400 TCTCGAGTCCTAGAGTTCAAGACCAGCCTAGGCAATGGTGAACCCCATCGGTACAAA 459

QY 2158 gaagccagcctgtagcctgaaagccttcttggagggcaggtctttccc-----c 2208

Db 460 AAATACAAAAATTAGCCAGGCATGGTGGCTGCGCATGTAGTCCAGCTCTCTGGGAGGC 519

QY 2209 aaaaatgcagacacgaagtgcaaaagtgaagctgcagctcttgcacaaagatgtaacttttc 2268

Db 520 TGAGTGGGAGGATCACTTGAACCCAGGACAGAGAGTTCAGATGAACCCGAGATCACGCC 579

QY 2269 acgaagccacagtggtgagggagagctgtcccacatttgcggaagtggctat----- 2321

Db 580 ACCAGCTCCAGCCTGGGCAACAGACAAAGACTCTGTCTAAAAAATACAAATAAAATAA 639

QY 2322 -----gtgagggacgggggaggtcccttagagataagagacaatcataagggagat 2376

Db 640 AAGTAGTCTTCACAGTACCAGCATTTCAATTTTCAAAAGATATAGAGCTAAAAAGGAAGGA 699

QY 2377 atcagagaaaaatcgtaaggggagcagatggtgtcaagagagaataggctgaccatcgagg 2436

Db 700 AAAAAAAGTAATGTTGGGCTTTTAAATACTCGTCTTAAATGTTCTTAAATGTTCTTAGGAGTG 759

QY 2437 actggcagaagctttca-----gaaacccactggacggctggcgacagtggc 2483

Db 760 CTGGGTTTTATTGTCAATCATTTATCTCTTTTAAAAATGTTATTGGCCAGGCACGGTGCC 819

QY 2484 ttaggcctgtaatcccgacacttttgggaggtgcgcaggtgaatcacttgaggtcagga 2543

Db 820 TCATGGCTGTAAATCCAGCACTTTGGGAGGCCGAGCGAGGAGGATCACCTGAGGTCAGGA 879

QY 2544 gttccagaccagcctggcccaacatggtgaaaccccatctctacagaaaaataaaatta 2603

Db 880 GTGTGAGACCAGCCTGGCCAAACATGGCGAAACCTG-TCTCTACTAAAAAATACAAAAATTA 938

QY 2604 gccaggcggtgtggcacaagcctagaatcccagctacttgggaggtgagc 2654

Db 939 ACTAGGCGTGTGTGTACGCTGTAGTCCAGCTACTCGGGAGGCTGAGG 989

Search completed: January 11, 2002, 12:14:02

Job time: 11194 sec

THIS PAGE BLANK (USPTO)

Parent Case # 101767341

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 11, 2002, 17:09:37 ; Search time 814.02 Seconds
(without alignments)
2795.191 Million cell updates/sec

Title: US-09-820-005-3_COPY_16200_18853
Perfect score: 2654
Sequence: 1 tgtctaggccatgattgc.....agctacttggagggtgagg 2654

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_1101.*
1: /SIDS8/gcgdata/geneseq/geneseq/NA1980.DAT.*
2: /SIDS8/gcgdata/geneseq/geneseq/NA1981.DAT.*
3: /SIDS8/gcgdata/geneseq/geneseq/NA1982.DAT.*
4: /SIDS8/gcgdata/geneseq/geneseq/NA1983.DAT.*
5: /SIDS8/gcgdata/geneseq/geneseq/NA1984.DAT.*
6: /SIDS8/gcgdata/geneseq/geneseq/NA1985.DAT.*
7: /SIDS8/gcgdata/geneseq/geneseq/NA1986.DAT.*
8: /SIDS8/gcgdata/geneseq/geneseq/NA1987.DAT.*
9: /SIDS8/gcgdata/geneseq/geneseq/NA1988.DAT.*
10: /SIDS8/gcgdata/geneseq/geneseq/NA1989.DAT.*
11: /SIDS8/gcgdata/geneseq/geneseq/NA1990.DAT.*
12: /SIDS8/gcgdata/geneseq/geneseq/NA1991.DAT.*
13: /SIDS8/gcgdata/geneseq/geneseq/NA1992.DAT.*
14: /SIDS8/gcgdata/geneseq/geneseq/NA1993.DAT.*
15: /SIDS8/gcgdata/geneseq/geneseq/NA1994.DAT.*
16: /SIDS8/gcgdata/geneseq/geneseq/NA1995.DAT.*
17: /SIDS8/gcgdata/geneseq/geneseq/NA1996.DAT.*
18: /SIDS8/gcgdata/geneseq/geneseq/NA1997.DAT.*
19: /SIDS8/gcgdata/geneseq/geneseq/NA1998.DAT.*
20: /SIDS8/gcgdata/geneseq/geneseq/NA1999.DAT.*
21: /SIDS8/gcgdata/geneseq/geneseq/NA2000.DAT.*
22: /SIDS8/gcgdata/geneseq/geneseq/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	204	7.7	2713	20	AA27233
2	197.6	7.4	3621	22	AA500824
3	196	7.4	32204	22	AA157790
4	195.6	7.4	32152	22	AA157791
5	192.4	7.2	2498	22	AA29453
6	189.8	7.2	160552	22	AA02697
7	189.6	7.1	32199	22	AA157673
8	189.4	7.1	2839	21	AA38992
9	189.4	7.1	2839	22	AAH74642
10	189.4	7.1	2839	22	AA02498
11	186	7.0	41684	21	AAA28150

c 12	185.8	7.0	6711	21	AAF21122	Human low adenosin
c 13	185.8	7.0	6711	21	AAA35000	Human adenosine re
c 14	185.8	7.0	10032	21	AAF21124	Human low adenosin
c 15	185.8	7.0	10032	21	AAA35002	Human adenosine re
c 16	184.2	6.9	2446	22	AAH18305	Human cDNA sequenc
c 17	182.2	6.9	121162	21	AAH66548	Human kinesin-like
c 18	181.6	6.8	56583	21	AAF21125	Human low adenosin
c 19	181.6	6.8	56583	21	AAA35003	Human adenosine re
c 20	181	6.8	43069	21	AA236335	Genomic sequence o
c 21	180.4	6.8	23187	21	AA50273	Human lipolysis st
c 22	180.4	6.8	23187	22	AA62331	Human leptin fragm
c 23	180	6.8	2001	22	AAH18441	Human cDNA sequenc
c 24	180	6.8	160552	22	AA02697	Human glycosyl sul
c 25	179	6.7	22976	20	AAH34326	Genomic region con
c 26	178.2	6.7	31853	22	AA163343	Human kidney relat
c 27	177.8	6.7	17431	22	AA157710	Human colorectal c
c 28	177	6.7	11967	22	AA97863	Human neuroblastom
c 29	177	6.7	21721	20	AAH34427	Human lipolysis st
c 30	176.8	6.7	13927	22	AA97853	Human neuroblastom
c 31	176.8	6.7	57728	22	AA87588	Human 9p11 chromos
c 32	176.8	6.7	121162	21	AA66548	Human kinesin-like
c 33	176	6.6	84607	20	AA90847	Human PACAP genom
c 34	174.2	6.6	78925	21	AA89888	Human FN gene. Ho
c 35	174	6.6	6405	22	AA97850	Human neuroblastom
c 36	173.2	6.5	110000	22	AA84800	Nucleotide sequenc
c 37	173	6.5	14426	22	AA162921	Human genomic DNA
c 38	173	6.5	26928	20	AA232184	Human prothrombin
c 39	172.6	6.5	32134	22	AA163522	Human kidney relat
c 40	172.6	6.5	32192	22	AA163523	Human kidney relat
c 41	172.2	6.5	32249	22	AA162932	Human genomic DNA
c 42	172.2	6.5	53526	19	AA194101	Human PKD1 gene.
c 43	172.2	6.5	53577	17	AA18551	Human polycystic k
c 44	172.2	6.5	53577	19	AA194108	Human PKD1 locus b
c 45	171.4	6.5	5197	22	AA162920	Human genomic DNA

ALIGNMENTS

RESULT 1
AA27233
ID AAX27233 standard; DNA; 2713 BP.
XX.
AC AAX27233;
XX
XX
DT 28-MAY-1999 (first entry)
XX
DE Human Rab protein, RABP-3, coding sequence.
XX

XX Rab protein; RABP-1; RABP-2; RABP-3; human; vesicle trafficking; cancer;
XX cell differentiation; apoptosis; immunodeficiency; cell proliferation;
XX neurodegenerative disease; myelodysplastic syndrome; wasting disease;
XX toxin-induced disease; infection; genetic defect; diagnosis; therapy; ss.

OS Homo sapiens.

PN WO9909182-A2.

XX 25-FEB-1999.

PD 17-AUG-1998; 98WO-US16983.

XX 21-AUG-1997; 97US-0916901.

XX (INCY-) INCYTE PHARM INC.

XX Corley NC, Hillman JL, Lal P, Shah P;

XX WPI; 1999-181042/15.

XX P-PSDB; AAY00920.

XX New purified human Rab proteins - used to develop products for

XX treating e.g. AIDS, immunodeficiencies, neurodegenerative diseases,

PT

Death-associated protein 6; DAXX; polymorphism; haplotype pair; human; immune disorder; autoimmune disease; population diversity; ds; paternity testing; anthropological lineage; forensic application.

```
Oy 1970 ccgagacggccgactgcactccagcctgggactacagagagagactctatctcaaaaaa 2029
Db 23894 CCAAGATCGCACCACTACACTCCAGCATGGCGCAACAGAGAGAGACTGTCTCAAAAAA 23835
Oy 2030 aaaaaaa-aaagtaactttaggtgcaggggtgctctgttattcactgagacccgtg 2087
Db 23834 TAAATAAATAGCGCAGGTGCAGTGGCTCATGCTGTAAATTCAGCAGCATTTGGGAGGCCGAG 23775
Oy 2088 cccgggttatgaggtgtaccagaaaagcaagtattcactatgcacactattcacgcgtca 2147
Db 23774 CGAGGCAGATCAGAAAGTCTAGGAGTTCAGACAGCCTGCCCAATATGGTGACACCCCC 23715
Oy 2148 ccctagcattgaagccagcctgtagcctgaaagcccttg-----ctttgagggcaggt 2200
Db 23714 GTCTCTACTAAAAATACAAAAATTAGCTGGCGGTGGTGGCTCGCGCCTTTAGTCCCACT 23655
Oy 2201 ctttcccaaaatcagacagaaaggtgcaaaagtgaagctccagctcttgcacaaagtgt 2260
Db 23654 ACTTGGGAAGCTGAGGCAGAGAAATCGCTTGAACCCAGGAGCGGAGATTTTCAGTGAGCC 23595
Oy 2261 aactgtcacgaagccacagagtgagcagggagagctgtccccacatttgcggaagtggcta 2320
Db 23594 AAGATGGTGGCCACTG-CACTCCAGCCTGGGTGACAGACCAAGACTCTCTCAAAAAATA 23536
Oy 2321 tgtgaggaacggggagggcggttcccttagagataagagacacaatcataaggggagatatca 2380
Db 23535 AATAAATAAATAAATATGCGCCAGCTGTGGCTCACCCTTATAATCCTAGCACTTTGGG 23476
Oy 2381 gaaaaaatcgaagggagagagatggtgtcaagagaaataggtgacctgaaggaactg 2440
Db 23475 AAGTTGA--GGTGGGTGGATTGCTGTGATCTCAGGATTACAGACCAAGCTAGGCAACATAG 23418
Oy 2441 gcagaagcttcc---agaaaaccactgagcggctgggcacagtggtccttagcctgtaact 2497
Db 23417 TGAGACCTCATCTCAATAAATCAATAAATAGCTGGGCGCACTGGCTCATGCCGTGTAATC 23358
Oy 2498 ccagcactttggaggctgacgcaggtgaaatcaccttaggtcaggagttccagaccagcc 2557
Db 23357 CCAGCACTTTGGAGGCTGAGGTGGCGGATCACTTGAGGTGAGGATTCAGACCAAGCC 23298
Oy 2558 tggcaacatggtgaaccccatctctacagaaaataaaaaattagccagggctggtgg 2617
Db 23297 TTGCCAATGATGAACCTTGTCTCTACTAAAAATACAAAAATTAGCTGGGCACGGTGG 23238
Oy 2618 cacaagctagaatccagctacttctgagggctgagg 2654
Db 23237 CACACACCTGTAGTCCCACTATTTGGGGCCTGAGG 23201

RESULT 3
AAI57790
ID AAI57790 standard; DNA; 32204 BP.
XX
XX
AC AAI57790;
XX
XX 19-OCT-2001 (first entry)
XX
XX Human colorectal cancer antigen coding sequence SEQ ID NO: 327.
XX
XX Human; colorectal cancer; colorectal cancer antigen; gene therapy; ds.
XX
XX Homo sapiens.
XX
XX WO200155350-A1.
XX
XX 02-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US01350.
XX
XX 31-JAN-2000; 2000US-0179065.
XX
XX 04-FEB-2000; 2000US-0180628.
XX
XX 24-FEB-2000; 2000US-0184664.
XX
XX 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
```


13-OCT-2000; 2000US-0239935.
 13-OCT-2000; 2000US-0239937.
 20-OCT-2000; 2000US-0240960.
 20-OCT-2000; 2000US-0241221.
 20-OCT-2000; 2000US-0241785.
 20-OCT-2000; 2000US-0241786.
 20-OCT-2000; 2000US-0241787.
 20-OCT-2000; 2000US-0241808.
 20-OCT-2000; 2000US-0241809.
 20-OCT-2000; 2000US-0241826.
 01-NOV-2000; 2000US-0244617.
 08-NOV-2000; 2000US-0246474.
 08-NOV-2000; 2000US-0246475.
 08-NOV-2000; 2000US-0246476.
 08-NOV-2000; 2000US-0246477.
 08-NOV-2000; 2000US-0246478.
 08-NOV-2000; 2000US-0246523.
 08-NOV-2000; 2000US-0246524.
 08-NOV-2000; 2000US-0246525.
 08-NOV-2000; 2000US-0246526.
 08-NOV-2000; 2000US-0246527.
 08-NOV-2000; 2000US-0246528.
 08-NOV-2000; 2000US-0246532.
 08-NOV-2000; 2000US-0246609.
 08-NOV-2000; 2000US-0246610.
 08-NOV-2000; 2000US-0246611.
 08-NOV-2000; 2000US-0246613.
 17-NOV-2000; 2000US-0249207.
 17-NOV-2000; 2000US-0249208.
 17-NOV-2000; 2000US-0249209.
 17-NOV-2000; 2000US-0249210.
 17-NOV-2000; 2000US-0249211.
 17-NOV-2000; 2000US-0249212.
 17-NOV-2000; 2000US-0249213.
 17-NOV-2000; 2000US-0249214.
 17-NOV-2000; 2000US-0249215.
 17-NOV-2000; 2000US-0249216.
 17-NOV-2000; 2000US-0249217.
 17-NOV-2000; 2000US-0249218.
 17-NOV-2000; 2000US-0249244.
 17-NOV-2000; 2000US-0249245.
 17-NOV-2000; 2000US-0249264.
 17-NOV-2000; 2000US-0249265.
 17-NOV-2000; 2000US-0249297.
 17-NOV-2000; 2000US-0249299.
 01-DEC-2000; 2000US-0250160.
 01-DEC-2000; 2000US-0250391.
 05-DEC-2000; 2000US-0251030.
 05-DEC-2000; 2000US-0251988.
 05-DEC-2000; 2000US-0256719.
 06-DEC-2000; 2000US-0251479.
 08-DEC-2000; 2000US-0251856.
 08-DEC-2000; 2000US-0251868.
 08-DEC-2000; 2000US-0251869.
 08-DEC-2000; 2000US-0251989.
 08-DEC-2000; 2000US-0251990.
 11-DEC-2000; 2000US-0254097.
 05-JAN-2001; 2001US-0259678.
 (HUMA-) HUMAN GENOME SCI INC.
 Rosen CA, Barash SC, Ruben SM;
 WPI; 2001-457727/49.
 Isolated polypeptide for treating, preventing and/ or prognosing
 disorders related to the colon and rectum including colorectal cancers
 and also for testing and detection e.g. diagnosis -
 Disclosure; SEQ ID NO: 327; 522pp + Sequence Listing; English.
 The present invention provides the protein and coding sequences of a

```
Db 13779 cagcactatgaagagctgaggcaggagagattgcttaagccctagagttcaagaccagact 13858
Qy 2151 tagcattgaagccagcctgtagcctgaaagccttctgttgaggcaggcttttcccccaa 2210
Db 13839 gggcaatgtagtga-----accccccatctctacaggaacaataatacaataaaa 13893
Qy 2211 aatgcagacacgaaggtgcaagtgaagctgccagctcttgcaaaagatgtaacttgtcac 2270
Db 13894 atggcagttctcagtttgcgtgctggttctatgtgttcacgtctcttcagaatacac 13953
Qy 2271 -gaagggcacagtgaggcaggagagctgtccacatttgcgggaagtggtctatgtgagac 2329
Db 13954 tgaatggccaggttcaactctcaagactggaagcagcagcaagagatgggatttceag 14013
Qy 2330 gggggaggcgggtcccttagagataagagacaatacataaaggggagatatcagagaaaac 2389
Db 14014 gccagaactccaggtatttttaatttttttaatttttaatttttaattttatacaggggccaagcac 14073
Qy 2390 gtaaggggagcagatggtgttcaagagagaataggctgaccatcgaaaggactggcagaagct 2449
Db 14074 ggt-ggctcacacctgtaattccggtactttgggagaaacatggcaaaacctgtctctac 14132
Qy 2450 ttcagaaaaccactggaagcgtggcacagtggttagcctgttaatccccagcactttgg 2509
Db 14133 taadaatacaaaaaatcaggctgdcgcgttagctcatgctgttaatccccagcactttgg 14192
Qy 2510 gaggtgacgcaggtgaaatcacttgaggtcagagagttccagaccagcctggccaacatgg 2569
Db 14193 gagccaaggcaggtggatcacctgagatcaggagtcgtgagccagcctggccaacgctgg 14252
Qy 2570 tgaacccccctctacagaaaataataaaaaattagccaggcgtggtggcacaagcctaga 2629
Db 14253 cgaacccccctctactaaaaatacaaaaaattagccgggcatggtgtgtacaccttta 14312
Qy 2630 atcccagctacttggggagctgaggg 2654
Db 14313 atcccagctactaaggagctgaggg 14337

RESULT 4
AAI57791
ID AAI57791 standard; DNA; 32152 BP.
AC AAI57791;
XX
XX
DT 19-OCT-2001 (first entry)
DE Human colorectal cancer antigen coding sequence SEQ ID NO: 328.
KW Human; colorectal cancer; colorectal cancer antigen; gene therapy; ds.
OS Homo sapiens.
XX
XX
PN WO200155350-A1.
XX
XX
PD 02-AUG-2001.
XX
XX
XX 17-JAN-2001; 2001WO-US01350.
XX
XX
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
```


RESULT	6
AAAD02697	AAAD02697 standard; DNA; 160552 BP.
AAAD02697	AAAD02697; (first entry)
02-MAY-2001	Human glycosyl sulfotransferase-4 (GST-4) genomic DNA.
DE	Human; glycosyl sulfotransferase-4; GST-4; immunosuppressive;
XX	therapy; selectin binding inhibitor; gene therapy; inflammation;
XX	systemic lupus erythematosus; SLE; rheumatoid arthritis; diabetes;
XX	polyarthritis nodosa; polymyositis; systemic sclerosis; dermatitis;
XX	glomerulonephritis; myasthenia gravis; Sjogren's syndrome; adenitis;
XX	Hashimoto's disease; Grave's disease; hypoparathyroidism; anaemia;
XX	demyelinating disease; cirrhosis; ulcerative colitis; allergic rhinitis;
XX	myocarditis; adult respiratory distress syndrome; eczema; psoriasis;
XX	asthma; hypersensitivity; rheumatic fever; tissue rejection;
XX	chromosome 16q23.1; ds.
OS	Homo sapiens.
XX	Key
XX	Location/Qualifiers
XX	exon
FT	32847..32922
FT	/*tag= a
FT	/number= 1
FT	/label= 4a_504
FT	32923..35592
FT	/*tag= b
FT	/cons_splice= (5'site:NO, 3'site:YES)
FT	35593..35674
FT	/*tag= c
FT	/number= 2
FT	/label= 4a_503
FT	35675..45093
FT	/*tag= d
FT	45094..45185
FT	/*tag= e
FT	/number= 3
FT	/label= 4a_502
FT	45186..46633
FT	/*tag= f
FT	/cons_splice= (5'site:NO, 3'site:NO)
FT	46634..46700
FT	/*tag= g
FT	/number= 4
FT	/label= 4a_501
FT	46701..47938
FT	/*tag= h
FT	/cons_splice= (5'site:YES, 3'site:NO)
FT	47939..49746
FT	/*tag= i
FT	/number= 5
FT	/note= "Includes 17 base pairs of 5'UTR, the ORF
FT	and all of 3'UTR"
FT	47939..47955
FT	/*tag= j
FT	/note= "Portion of 5' untranslated region (5'UTR)"
FT	47956..49128
FT	/*tag= k
FT	/product= "Human glycosyl transferase-4alpha
FT	(GST-4alpha)"
FT	49129..49746
FT	/*tag= l
FT	83257..83347
FT	/*tag= m
FT	/label= 4a_502
FT	83348..96412
FT	/*tag= n
FT	/cons_splice= (5'site:NO, 3'site:NO)
FT	96413..96484
FT	/*tag= o
FT	intron
FT	96485..98456
FT	/*tag= p
FT	/cons_splice= (5'site:NO, 3'site:NO)
FT	98457..99968
FT	/*tag= q
FT	/note= "Includes 17 base pairs of 5'UTR, the ORF
FT	and all of 3'UTR"
FT	98457..98473
FT	/*tag= r
FT	/note= "Portion of 5' untranslated region (5'UTR)"
FT	98474..99661
FT	/*tag= s
FT	/product= "Human glycosyl transferase-4beta
FT	(GST-4beta)"
FT	99662..99968
FT	/*tag= t
XX	WO200106015-A1.
XX	25-JAN-2001.
XX	19-JUL-2000; 2000WO-US19741.
XX	20-JUL-1999; 99US-0144694.
XX	13-JUL-2000; 2000US-0593828.
XX	(RECC) UNIV CALIFORNIA.
XX	Rosen SD, Lee JK, Hemmerich S;
XX	WPI: 2001-138471/14.
XX	P-PSDB; AAY72639, AAY72640.
XX	New glycosyl sulfotransferases (GST)-4alpha, GST-4beta and GST-6 for
XX	diagnostic and therapeutic agent screening applications .
XX	Example 1; Page 62-104; 128pp; English.
XX	The present sequence is human glycosyl sulfotransferase-4 (GST-4) genomic
XX	DNA encoding GST-4alpha and GST-4beta. GST-4 gene is found on
XX	chromosome 16q23.1.
XX	GST is a type 2 membrane protein useful for inhibiting a binding event
XX	between a selectin and a selectin ligand, which comprises contacting the
XX	selectin with a non-sulphated selectin ligand, GST and a small molecular
XX	agent that inhibits the sulphation activity of GST. GST is also useful
XX	in inhibiting a selectin mediated binding event. GST is useful in gene
XX	therapy to treat disorders such as acute or chronic inflammation,
XX	systemic lupus erythematosus (SLE), rheumatoid arthritis, polyarthritis
XX	nodosa, polymyositis, dermatomyositis, systemic sclerosis, diabetes,
XX	glomerulonephritis, myasthenia gravis, Sjogren's syndrome, Hashimoto's
XX	disease, Grave's disease, adenitis, hypoparathyroidism, pernicious
XX	anaemia, demyelinating diseases, cirrhosis, ulcerative colitis,
XX	dermatitis, myocarditis, regional enteritis, adult respiratory distress
XX	syndrome, infantile eczema, psoriasis lichen planus, allergic rhinitis,
XX	bronchial asthma, hypersensitivity, rheumatic fever and tissue rejection
XX	during transplantation.
XX	Sequence 160552 BP; 40281 A; 37573 C; 38015 G; 44564 T; 119 other:
XX	Query Match
XX	Best Local Similarity 7.2%; Score 189.8; DB 22; Length 160552;
XX	Matches 451; Conservative 0; Mismatches 287; Indels 49; Gaps 4;
XX	1869 aaaaattagctggcggtggcgatgccacacccagctactccagctggaggctgagc 1928
XX	
XX	54991 aaaaattagctggcggtggcgatgccacacccagctactccagctggaggctgagc 55050
XX	
XX	1929 atgagaatcgtctgaacccggggaggcagatgttcagtgagccgagcgcacactgca 1988
XX	
XX	55051 aggaagatcacatcaacctggagggcagagggctgcaagctctagatggatgctgca 55110
XX	


```

FT FT /*tag= x
FT FT /number= 12
FT FT intron 28829..34698
FT FT /*tag= y
FT FT /number= 12
FT FT exon 34699..34791
FT FT /*tag= z
FT FT /number= 13
FT FT intron 34792..36678
FT FT /*tag= aa
FT FT /number= 13
FT FT exon 36679..36361
FT FT /*tag= ab
FT FT /number= 14
FT FT intron 36862..39013
FT FT /*tag= ac
FT FT /number= 14
FT FT exon 39014..39169
FT FT /*tag= ad
FT FT /number= 15
FT FT intron 39170..39455
FT FT /*tag= ae
FT FT /number= 15
FT FT exon 39456..39684
FT FT /*tag= af
FT FT /number= 16
FT FT variation replace(6491,T)
FT FT /*tag= ag
FT FT variation replace(15234,G)
FT FT /*tag= ah
FT FT variation replace(15868,T)
FT FT /*tag= ai
FT FT variation replace(16729,T)
FT FT /*tag= aj
FT FT variation replace(18311,C)
FT FT /*tag= ak
FT FT variation replace(18572,T)
FT FT /*tag= al
FT FT variation replace(22906,A)
FT FT /*tag= am
FT FT variation replace(23175,T)
FT FT /*tag= an
FT FT variation replace(23253,G)
FT FT /*tag= ao
FT FT variation replace(26106,C)
FT FT /*tag= ap
FT FT variation replace(30464,G)
FT FT /*tag= aq
FT FT variation replace(30669,C)
FT FT /*tag= ar
FT FT variation replace(31250,A)
FT FT /*tag= as
FT FT variation replace(35148,G)
FT FT /*tag= at
FT FT variation replace(36801,G)
FT FT /*tag= au
FT FT variation replace(37286,T)
FT FT /*tag= av
FT FT variation replace(37536,C)
FT FT /*tag= aw
FT FT variation replace(39321,A)
FT FT /*tag= ax
FT FT variation replace(39689,T)
FT FT /*tag= ay
FT FT primer_bind 6137..6157
FT FT /*tag= az
FT FT primer_bind complement (6577..6597)
FT FT /*tag= ba
FT FT primer_bind 14864..14882
FT FT /*tag= bb
FT FT primer_bind complement (15292..15312)
FT FT /*tag= bc
FT FT primer_bind 15837..15855

```

```

FT FT /*tag= bd
FT FT complement (16429..16266)
FT FT /*tag= be
FT FT primer_bind 16599..16617
FT FT /*tag= bf
FT FT complement (17030..17049)
FT FT /*tag= bg
FT FT primer_bind 18131..18150
FT FT /*tag= bh
FT FT complement (18592..18610)
FT FT /*tag= bi
FT FT primer_bind 22710..22727
FT FT /*tag= bj
FT FT primer_bind 23100..23118
FT FT /*tag= bk
FT FT complement (23130..23149)
FT FT /*tag= bl
FT FT complement (23512..23530)
FT FT /*tag= bm
FT FT primer_bind 25822..25840
FT FT /*tag= bn
FT FT primer_bind complement (26222..26241)
FT FT /*tag= bo
FT FT primer_bind 30332..30352
FT FT /*tag= bp
FT FT primer_bind complement (30754..30773)
FT FT /*tag= bq
FT FT primer_bind 30918..30935
FT FT /*tag= br
FT FT complement (31390..31408)
FT FT /*tag= bs
FT FT primer_bind 34780..34799
FT FT /*tag= bt
FT FT complement (35215..35233)
FT FT /*tag= bu
FT FT primer_bind 36593..36610
FT FT /*tag= bv
FT FT primer_bind complement (37017..37036)
FT FT /*tag= bw
FT FT primer_bind 37060..37080
FT FT /*tag= bx
FT FT primer_bind complement (37541..37561)
FT FT /*tag= by

```

```

Query Match 7.0% Score 186; DB 21; Length 41684;
Best Local Similarity 56.8% Pred. No. 5.7e-28;
Matches 452; Conservative 0; Mismatches 320; Indels 24; Gaps 5;

Qy 1869 aaaaattagctggcggtggtgcatccacatccacgactgctggtggaggtgagggc 1928
Db 7592 AAAAAATTAGCGGAGCGTGTGCGGGGCGCACCTGTATCCAGCTACTGGGAGGCTGAGGC 7533

Qy 1929 atgagaatcgcttgaacccggggagcgagatgttcagtgagccgagagcgccactgca 1988
Db 7532 AGGAGAATCACCTTGAACCCCAAGAGTTGGAGGTGCGAGTGCAGTGCAGATCATGCCACTGCA 7473

Qy 1989 ctccagctgactacagagcgagactctatctcaaaaaaiaaaaaa-----aa 2039
Db 7472 CTCAGCGCTGGCGGACAGAGCAAGACTCCATCTCAAAAAACAAACAAACACATGAA 7413

Qy 2040 aaaagtaacttagtgaggggtgctctctgttattcactgagacggtgccccggttga 2099
Db 7412 GATGCATATCAAGTGAAGACATAAAAGACTGTTTAAATAAGTGGTGTGCTATTG 7353

Qy 2100 ggttgtaaccagaagcaagtatttcactatgcacactatccaccgtccaccctagcattga 2159
Db 7352 GAATAACGGGTTAAACCCCAAAATAATATTCCTTATGAATCAAAAGGTTTAGCATTTA 7293

Qy 2160 agccagctgtagcctaagaagccttctgtgagggcaggtcttcccccaaaatgcacac 2219
Db 7292 AAAAATCATGCATGTAGCAA-AAITAAATATGTGAGGAATGCTTTTCCAAAGTTTAAACAA 7234

Qy 2220 acgaagggtcaaggtgaagctgcccagttctgtcaaaagatgttaacttgtcaagagggccac 2279

```

Db 7233 AAGATTGAGAGTTCTGATTACAGAAATCTTAAATAAAAGTTTACATGGGAAAAAAG 7174
 Qy 2280 gagtggcaggagagctgtccacatttgcggaagtggctatgtgagagcggggagcgcg 2339
 Db 7173 TAAACTAGAAAAATAATACACAAATGTGACAAAGGGTCAAGTGAAGGACAGTAAATATT 7114
 Qy 2340 ggtcccttagagataaagagacaatacataagggggagagatcacagaaatcgtaaagggag 2399
 Db 7113 AATCCATATCAACAGCAGACCCCTAGNAAGTCATAACCCAGACACTCTCGCA----- 7058
 Qy 2400 cagatggtttcaagagagaataggctgaccatcgaaggactggcagaagccttccagaaac 2459
 Db 7059 -----GACTTTATAAGTTGATGTTGTTGTAATAAATGCCACCAATATATAAGAAAAAT 7006
 Qy 2460 cactgagcgtgggcacagctggttagcctgtaatacccagcactttggagagctgacg 2519
 Db 7005 AAAGCAGGGCCAGGTGGGGTGCACACCTGTAAATCCAGCACTTTGGAGGCCAAGG 6946
 Qy 2520 caggtgaatcacttgaggtca-ggagttccagacagcctggcccaacatggtgaacccc 2578
 Db 6945 CGAATGGATCACCCTGAGGTAGGAGTTCAGACAGCCTGACCAACATGTGTAACCCC 6886
 Qy 2579 atctctacagaaaataaaaaattagccagcgtggtggcgaagcctagatccccagct 2638
 Db 6885 GCCTCTACTAAAAATA-CAAAAATTAGCCAGGCATGGTGGGCGCATGCCTATAATCCAGCT 6827
 Qy 2639 acttggagagctgagg 2654
 Db 6826 ACTTGGGAGGCTGAGG 6811

RESULT 12

AAE21122/c
 ID AAE21122 standard; DNA: 6711 BP.

XX AAE21122;
 AC
 XX
 XX
 DT 14-MAR-2001 (first entry)
 XX
 XX Human low adenosine antisense oligonucleotide related sequence #2689.
 DE
 XX Low adenosine antisense oligonucleotide; phosphorothioate; allergy;
 KW human; airway disorder; bronchoconstriction; lung inflammation;
 KW surfactant depletion; respiratory; bronchodilator; antiinflammatory;
 KW immunosuppressive; antialsthmatic; analgesic; hypotensive; cytostatic;
 KW respiratory obstruction; pulmonary obstruction; impeded respiration;
 KW surfactant hypoproduction; pulmonary vasoconstriction; asthma; RDS;
 KW respiratory distress syndrome; pain; cystic fibrosis; allergic rhinitis;
 KW pulmonary hypertension; emphysema; pulmonary transplantation rejection;
 KW chronic obstructive pulmonary disease; pulmonary infection; bronchitis;
 KW cancer; ss.
 XX
 OS Homo sapiens.
 XX
 XX W020062736-A2.
 PN
 XX
 XX 26-OCT-2000.
 XX
 XX 24-MAR-2000; 2000WO-US08020.
 PF
 XX
 XX 06-APR-1999; 99US-0127958.
 PR
 XX (UYEC-) UNIV EAST CAROLINA.
 PA
 PA (NYCE/) NYCE J W.
 XX
 XX Nyce JW;
 PI
 XX WPI; 2000-679539/66.
 DR
 XX
 XX Low adenosine (A) content antisense oligonucleotides which do not
 PT trigger adenosine receptors during metabolism, useful e.g. for treating
 PT cancers and respiratory obstructions -

XX
 PS
 XX
 CC The present invention describes low adenosine (A) content antisense
 CC oligonucleotides and compositions (I) comprising them. In the antisense
 CC oligonucleotides the A is replaced by a 'Universal' or alternative base.
 CC (1) can have respiratory, bronchodilator, antiinflammatory, analgesic,
 CC immunosuppressive, antialsthmatic, hypotensive and cytostatic activities.
 CC The antisense oligonucleotides and (I) can be used to down-regulate the
 CC expression and/or activity of target polypeptides associated with
 CC lung/respiratory disorders and malignancies, such as stimulating and
 CC activating peptide factors and transmitters, transcription factors,
 CC immunoglobulins and antibodies, antibody receptors, cytokines and
 CC chemokines, endogenously produced specific and non-specific enzymes,
 CC binding proteins, adhesion molecules and their receptors, cytokine and
 CC chemokine receptors, adenosine receptors, bradykinin receptors, central
 CC nervous system (CNS) and peripheral nervous and non-nervous system
 CC receptors, CNS and peripheral nervous and non-nervous system peptide
 CC transmitters, defensins, growth factors, vasoactive peptides and
 CC receptors, binding proteins and malignancy associated proteins. The
 CC antisense oligonucleotides may be used in this way to treat disorders
 CC including respiratory obstruction (especially pulmonary obstruction
 CC and/or bronchoconstriction) and/or lung inflammation, allergy(ies)
 CC and/or surfactant hypoproduction which are associated with a disease or
 CC condition selected from pulmonary vasoconstriction, inflammation,
 CC allergies, asthma, impeded respiration, respiratory distress syndrome
 CC (RDS), pain, cystic fibrosis (CF), allergic rhinitis (AR), pulmonary
 CC hypertension, emphysema, chronic obstructive pulmonary disease (COPD),
 CC pulmonary transplantation rejection, pulmonary infections, bronchitis,
 CC and/or cancer. AAF18434 to AAF21543 represent human polynucleotide
 CC fragments and antisense oligonucleotides used in the exemplification of
 CC the present invention.
 XX
 SQ Sequence 6711 BP; 1552 A; 1625 C; 1717 G; 1817 T; 0 other;

Query Match 7.0%; Score 185.8; DB 21; Length 6711;
 Best Local Similarity 55.7%; Pred. No. 4.8e-28;
 Matches 443; Conservative 0; Mismatches 337; Indels 15; Gaps 4;

Qy 1869 aaaaattagctggcgtggtgcatgccacaaatcccagctactctggggaggtgagggc 1928
 Db 5916 AAAAATTAGCCGGGCATGGTGGCATGTGCCATATAGTCCAGCTACTCGGGAGGCTGAGGC 5857
 Qy 1929 atgagaatcgttaaacgggagcagatcttcagtgagccgagcgcgcactgca 1988
 Db 5856 AGGAGAAATGGCGTGAACCTGGGAGCGGAGCTTCAGTGTAGCCGAGATCGGCCACTGCA 5797
 Qy 1989 ctccagcctgactacagagcagactctctctcaaaaaa-aaaaaataaaagtaa 2047
 Db 5796 CTCAGCGCTGGGCGACAGAGCGAAATCCATCTCAAAAAGAGTTACACATATAGAAATGA 5737
 Qy 2048 cttaggtgcagggtgtcctctgtattcactgagacccgtgccccgggttatgaggtgtac 2107
 Db 5736 ATTGCGCCCCCTCCCAAAAACAAAACAAAACAAAACAAAACAAAACAAAACAAAAC 5677
 Qy 2108 cagaaagcaagtatttcactatgcactattcccgctccacctagcattgaagccagcc 2167
 Db 5676 ACCCAACCAACCCGATGTTATTATAGTAAATAGAACAGCAGCACTTAATGGTGGGCAC 5617
 Qy 2168 ttagcctgaaagccttctgttggagggcaggtcttcccccaaaatgcagacacagaaggt 2227
 Db 5616 AGTGGCTCATGCCGTAAATCCAGCACTTCGGGGGGGTCAAGGCAGGCAGATCACTTGAG 5557
 Qy 2228 gcaagtgaaagctccagctgttgcataaagatgaactgtc-----acgaagggccac 2279
 Db 5556 CCCAGGAGTTCGAGACCCAGCCTGGACAACTGGTAATCCCGCTCTACTAATAATACACA 5497
 Qy 2280 gagtggcaggagagagctgtccacacatttgcggaagtgagagcggggagcgcg 2339
 Db 5496 ATTTAGCTGGGAGTGGTGCATGCCCTGTAGTTCAGCACTACTTGAG--AGGCTGAGGCA 5439
 Qy 2340 ggtcccttagagataaagagacaatcataaaggggagagatcatcagagaaaaatcgtaaagggag 2399

Db 5438 GAAGAAATTGCTTGAACCCGGGAGGAGGTTTCAGCTGAGCTGAGATGGCCACCACTGCCTCC 5379
QY 2400 cagatgtttcgaagaaatagctgaccatcgaagactggcagaagctttcagaaaac 2459
Db 5378 AACCTGGGAGATAGAG-----CGAGATCCAGTCTGAGAAAAGAAAAAAGG 5323
QY 2460 cactggacgctggcacagtgttagctgtgaatccacagcactttggagggctgacg 2519
Db 5322 ACTTAATGTGTGGCGCAGTGCCTCATGCTGAAATCCAGCAGCTTTGGGGGCGCAAGG 5263
QY 2520 cagtgtaacacttgaggtcagaggttccagaccagcagctggcgaacatggtgaaaccca 2579
Db 5262 CAGGTGTTTACCTTAAGTCACGAGTTCAAGACCAAGCTTGGCCAAACATGATGAACCCCG 5203
QY 2580 tctctcagaaaaataaaaattagccagcggtgtgtgacacagcctagatcccaagcta 2639
Db 5202 TCTCCACAAAAATACAAAAATTAGCCAGGATGCTGGCCACACATCTGTATTCTCAGCTA 5143
QY 2640 cttggggaggctgag 2654
Db 5142 CTCAGGAGGCCAAGG 5128
RESULT 13
ID AAA35000/c
AC AAA35000;
XX 28-JUL-2000 (first entry)
XX Human adenosine receptor related polynucleotide SEQ ID NO:2689.
XX Human; adenosine receptor; low adenosine antisense oligonucleotide;
KW phosphothioate; impaired respiration; inflammation; allergy;
KW allergic disease; bronchoconstriction; inhibitor; antiinflammatory;
KW antiallergic; antiasthmatic; cytosolic; analgesic; impaired airway;
KW lung disease; ischaemic condition; pulmonary vasoconstriction; asthma;
KW respiratory distress syndrome; pain; cystic fibrosis; emphysema;
KW pulmonary hypertension; chronic obstructive pulmonary disease; COPD;
KW cancer; leukaemia; lymphoma; carcinoma; metastasis; ss.
XX Homo sapiens.
XX OS
XX WO200009525-A2.
XX 24-FEB-2000.
XX 03-AUG-1999; 99WO-US17712.
XX 03-AUG-1998; 98US-0095212.
XX (UYEC-) UNIV EAST CAROLINA.
XX PA
XX NYce JW;
XX WPI; 2000-205971/i8.
XX New antisense oligonucleotides useful for treating e.g. pulmonary
PT vasoconstriction, inflammation, allergies, asthma, hypertension,
PT bronchitis, emphysema, respiratory distress syndrome, ischemia or
PT cancers -
XX
XX Disclosure: Page 898-900; 1343pp; English.
XX The present invention describes a new composition comprising an
CC antisense oligonucleotide (ON) with low adenosine (up to 15%), which
CC targets nucleic acids involved in bronchoconstriction, allergies, and/or
CC inflammation. The ON can have antiinflammatory, antiallergic,
CC antiasthmatic, cytosolic and analgesic activities. The compositions are
CC useful for the treatment of diseases associated with inflammation,
CC impaired airways, including lung disease and diseases whose secondary

CC effects afflict the lungs of a subject. They can be used for treating
CC e.g. ischaemic conditions, pulmonary vasoconstriction, allergies,
CC asthma, impaired respiration, respiratory distress syndrome, pain, cystic
CC fibrosis, pulmonary hypertension, emphysema, chronic obstructive
CC pulmonary disease (COPD), and cancers such as leukaemias, lymphomas,
CC carcinomas, and cancers which may metastasise to the lungs, including
CC breast and prostate cancer. The reduction of the adenosine content of
CC the ONs reduces side effects. The A-containing ONs break down with the
CC release of deoxyadenosine which activates adenosine receptors causing
CC bronchoconstriction and inflammation. AAA32313 to AAA3512 represent the
CC nucleotide sequences given in the sequence listing from the present
CC invention, which correspond to SEQ ID NO:1 to 2815, and then the last
CC 185 sequences are also called SEQ ID NO:1 to 185, but the sequences
CC differ from the previously named sequences. SEQ ID NO:11 to 1680
CC (AAA32323 to AAA33992) are specifically claimed ONs from the present
CC invention. N.B. Sequences given in the disclosure of the present
CC invention do not match up with their corresponding SEQ ID NO: sequences
CC given in the sequence listing.
XX
SQ Sequence 6711 BP; 1552 A; 1625 C; 1717 G; 1817 T; 0 other;
Query Match 7.0%; Score 185.8; DB 21; Length 6711;
Best Local Similarity 55.7%; Pred. No. 4.8e-28;
Matches 443; Conservative 0; Mismatches 337; Indels 15; Gaps 4;
QY 1869 aaaaattagctggcggtgtggcatgcataccacaatccagactactgggagggctgagggc 1928
Db 5916 AAAAAATTAGCGGGCATGTTGGCATGTGCTTATAGTCCAGCTACTCGGAGGCTGAGGC 5857
QY 1929 atgagaatcgtttgaaccgggggagcgagatgttgtagtgagcgagcgccgactgca 1988
Db 5856 AGGAGAATGCGTGAACCTGGGAGCGGAGCTTCAGCTGAGCGCCGAGATCGGCCACTGCA 5797
QY 1989 ctccagctggactacagagcgagactctctcaaaaa-aaaaaataaaaaaagtaa 2047
Db 5796 CTCAGGCTGGGCGACAGAGCGAACTCCATCTCAAAAAGAGTTACACATATAGAATGA 5737
QY 2048 cttagtgtaggggtgctctgttattcactgagacggtgcccgttatgaggtgtac 2107
Db 5736 ATTGCGGCCCTCTCCCAAAAACAAACAAACAAACAAACAAACAAACAAACAAACAA 5677
QY 2108 cagaaagcaagatttccactgcacactattccacgctccacccctagcattgaagccagcc 2167
Db 5676 ACCCAACCAACCGCATGTATTATATAGTAAACATAAAGCAGCAGCTTAATTGGTGGGCAC 5617
QY 2168 ttagcctgaaagccttctgttgagggcgaggtcttcccccaaaatgcagacacgaaggt 2227
Db 5616 AGTGGCTCATGCTTAAATCCAGCAGCTTCGGGGGGGTCAAGGAGGAGCATCATCTTGAG 5557
QY 2228 gcaagtgaaagctgcccagctcttcaaaaagatgtaactgtc-----acgaagggccac 2279
Db 5556 CCCAGGAGTTGGAGACCAAGCCTGGACACATGGTAAATFCCCGCTCTACTATATATACACA 5497
QY 2280 gagtggcagggagagctgtccacatttgcggaagtgagtgagcagcggggagggcg 2339
Db 5496 ATTTAGCTGGGAGTGGTGGTGCATGCCCTGTAGTTCAGCTACTTTGAG--AGGCTGAGCA 5439
QY 2340 ggtcccttagagataaagagacaatcataaggggagagatatcagagaaaatcgtlaagggag 2399
Db 5438 GAACAATTGCTTGAACCCGGGAGGAGGTTCAGTGCAGTGAGATGGCACCAGCTGCACCTCC 5379
QY 2400 cagatgtttcgaagaaatagctgaccatcgaagactggcagaagctttcagaaaac 2459
Db 5378 AACCTGGGAGATAGAG-----CGAGATCCAGCTCTGAGAAAAGAAAAAAGG 5323
QY 2460 cactggacgctggcacagtggttagcctgtaatccacagcactttggagggctgacg 2519
Db 5322 ACTTAATGTGTGGCGCAGTGGCTCATGCTGAAATCCAGCAGCTTTGGGGGCGCAAGG 5263
QY 2520 cagtgtaacacttgaggtcagaggttccagaccagcagctggcgaacatggtgaaaccca 2579
Db 5262 CAGGTGTTTACCTTAAGTCACGAGTTCAAGACCAAGCTTGGCCACACATCTGTATTCTCAGCTA 5203

Qy 2580 tctctacagaaataataaaattagccagcggtgtgacacagcctagaaatccagcta 2639
 |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
 Db 5202 TCTCCACAAAATACAAAATATTAGCCAGGATGCTGGCACACATCTCTATTCTCAGCTA 5143
 |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
 Qy 2640 cttgggagcgtgagg 2654
 || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
 Db 5142 CTCAGGAGGCCAAGG 5128
 || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||

RESULT 14

AAF21124/c
 ID AAF21124 standard; DNA; 10032 BP.

AC AAF21124;

XX 14-MAR-2001 (first entry)

XX Human low adenosine antisense oligonucleotide related sequence #2691.

DE Low adenosine antisense oligonucleotide; phosphorothioate; allergy;
 KW human; airway disorder; bronchoconstriction; lung inflammation;
 KW surfactant depletion; respiratory; bronchodilator; antiinflammatory;
 KW immunosuppressive; antiasthmatic; analgesic; hypotensive; cytostatic;
 KW respiratory obstruction; pulmonary obstruction; impeded respiration;
 KW surfactant hypoproduction; pulmonary vasoconstriction; asthma; RDS;
 KW respiratory distress syndrome; pain; cystic fibrosis; allergic rhinitis;
 KW pulmonary hypertension; emphysema; pulmonary transplantation rejection;
 KW chronic obstructive pulmonary disease; pulmonary infection; bronchitis;
 KW cancer; ss.

XX Homo sapiens.

OS WO200062736-A2.

PN 26-OCT-2000.

XX 24-MAR-2000; 2000WO-US08020.

XX 06-APR-1999; 99US-0127958.

XX (UYEC-) UNIV EAST CAROLINA.

PA (NYCE/) NYCE J W.

XX Nyce JW;

XX WPI; 2000-679539/66.

DR Low adenosine (A) content antisense oligonucleotides which do not
 XX trigger adenosine receptors during metabolism, useful e.g. for treating
 PT cancers and respiratory obstructions -

XX Disclosure; Page 975-978; 1592pp; English.

XX The present invention describes low adenosine (A) content antisense
 CC oligonucleotides and compositions (I) comprising them. In the antisense
 CC oligonucleotides the A is replaced by a 'Universal' or alternative base.
 CC (I) can have respiratory, bronchodilator, antiinflammatory, analgesic,
 CC immunosuppressive, antiasthmatic, hypotensive and cytostatic activities.
 CC The antisense oligonucleotides and (I) can be used to down-regulate the
 CC expression and/or activity of target polypeptides associated with
 CC lung/respiratory disorders and malignancies, such as stimulating and
 CC activating peptide factors and transmitters, transcription factors,
 CC immunoglobulins and antibodies, antibody receptors, cytokines and
 CC chemokines, endogenously produced specific and non-specific enzymes,
 CC binding proteins, adhesion molecules and their receptors, cytokine and
 CC chemokine receptors, adenosine receptors, bradykinin receptors, central
 CC nervous system (CNS) and peripheral nervous and non-nervous system
 CC receptors, CNS and peripheral nervous and non-nervous system peptide
 CC transmitters, defensins, growth factors, vasoactive peptides and
 CC receptors, binding proteins and malignancy associated proteins. The
 CC antisense oligonucleotides may be used in this way to treat disorders
 CC including respiratory obstruction (especially pulmonary obstruction

CC and/or bronchoconstriction) and/or lung inflammation, allergy(ies)
 CC and/or surfactant hypoproduction which are associated with a disease or
 CC condition selected from pulmonary vasoconstriction, inflammation,
 CC allergies, asthma, impeded respiration, respiratory distress syndrome
 CC (RDS), pain, cystic fibrosis (CF), allergic rhinitis (AR), pulmonary
 CC hypertension, emphysema, chronic obstructive pulmonary disease (COPD),
 CC pulmonary transplantation rejection, pulmonary infections, bronchitis,
 CC and/or cancer. AAF18434 to AAF21543 represent human polynucleotide
 CC fragments and antisense oligonucleotides used in the exemplification of
 CC the present invention.

XX
 SQ Sequence 10032 BP; 2342 A; 2486 C; 2643 G; 2561 T; 0 other;

Query Match 7.0%; Score 185.8; DB 21; Length 10032;
 Best Local Similarity 55.7%; Pred. No. 5.1e-28;
 Matches 443; Conservative 0; Mismatches 337; Indels 15; Gaps 4;

Qy 1869 aaaaattagctggcggtgtgcatccacacacccagctactggtggagcctgagggc 1928
 |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
 Db 8514 AAAAATTAGCCGGCATGCTGGCATGTGCTTATGTCCACCTACTCGGAGGCTGAGGC 8455
 |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
 Qy 1929 atgagaatcgcttgaaccggggagcgagatgttgcagtgagccgagcgccactgca 1988
 |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
 Db 8454 AGGAGAAATGGCGTGAACCTGGGAGCGGAGCTTGCAGTGAAGCCGACCTGCA 8395
 |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
 Qy 1989 ctcagcctgagactacagagcgagactctctcaaaaaa-aaaaaaagagtaa 2047
 |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
 Db 8394 CTCAGGCTGGGCGACAGAGCGAACTCCATCTCAAAAAGAGTTACACATATAGAATGA 8335
 |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
 Qy 2048 cttaggtgcagggtgtccctctgttattcaactgagacggtgcccggttatgaggtgtac 2107
 |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
 Db 8334 ATTGCGCCCCCTCCCAAAAACAACAAACAAACAAACAAACAAACAAACAAACAA 8275
 |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
 Qy 2108 cagaaagcaagtattoactatgcacactattcacgctcaccctagcattgaaagccagcc 2167
 |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
 Db 8274 ACCCAACCAACCCGATGTTATTAATAGTAAACATAAAAGCAGCAGCTTAATGGTGGGCAC 8215
 |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
 Qy 2168 tgtagcctgaaagccttgccttggagggcaggtctttcccaaaatcagacagcagaggt 2227
 |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
 Db 8214 AGTGCCTCATGCTGTAAATCCACACACTTCGGGGGGGTCAAGGCAGCAGATCATCTTGAG 8155
 |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
 Qy 2228 gcaagtgagcgtccagctcttgcaaaagatgtaactgtctc-----acgaagggcac 2279
 |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
 Db 8154 CCCAGGAGTTTCGAGACCAGCCTGGACAAACATGGTAAATCCCGCTCTACTAATAATACACA 8095
 |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
 Qy 2280 gagtggcaggagagcgtgtccacatttgcggaagtgctgtgtgagcagggggagggcg 2339
 |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
 Db 8094 ATTTAGCTGGGAGTGGTGGTGCATGCCCTGCTAGTTCCAGCTACTTTGAG--AGGCTGAGGCA 8037
 |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
 Qy 2340 ggtcccttagagataagagacaatcataaggggagatatcagagaaatcgttaaggggag 2399
 |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
 Db 8036 GAAGAAATGCTTGAACCCGGGAGGAGTTGCGATGAGCTGAGATGGCCACCCTCCTCCTCC 7977
 |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
 Qy 2400 cagatggtgtcagagaaataggtgctgacctcgaaggaactggcagaagccttcagagaac 2459
 |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
 Db 7976 AACCTGGGAGATAAG--CGAGATCCAGCTCTGAGAAAAGAAAAAAAAGC 7921
 |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
 Qy 2460 cactggcagcctggcagcagtgcttagcctgttaatccacagcacttgggagcgtgacg 2519
 |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
 Db 7920 ACTTAATTTGGTGGGCGCAGTGGCTCATGCTGAAATCCAGCAGCTTTGGGGGGCCCAAGG 7861
 |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
 Qy 2520 cagtgtaactcactgaggtcagaggttccagaccagcctggccaacatggtgaacccca 2579
 |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
 Db 7860 CAGTGGTTTCACTTAAGTACAGAGTTCAGACCAGCTTGGCCCAACATGATGAACCCCG 7801
 |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
 Qy 2580 tctctacagaaataataaaattagccagcggtgtgacacagcctagaaatccagcta 2639
 |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
 Db 7800 TCTCCACAAAATACAAAATATTAGCCAGGATGCTGGCACACATCTCTATTCTCAGCTA 7741
 |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
 Qy 2640 cttgggagcgtgagg 2654
 || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
 Db 7740 CTCAGGAGGCCAAGG 7726
 || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||

RESULT 15

AAA35002/c

ID AAA35002 standard; DNA; 10032 BP.

XX AC AAA35002;

XX 28-JUL-2000 (first entry)

XX DE Human adenosine receptor related polynucleotide SEQ ID NO:2691.

XX KW Human; adenosine receptor; low adenosine antisense oligonucleotide; phosphothioate; impaired respiration; inflammation; allergy; allergic disease; bronchoconstriction; inhibitor; antiinflammatory; antiallergic; antiasthmatic; cytosstatic; analgesic; impaired airway; lung disease; ischaemic condition; pulmonary vasoconstriction; asthma; respiratory distress syndrome; pain; cystic fibrosis; emphysema; pulmonary hypertension; chronic obstructive pulmonary disease; COPD; cancer; leukaemia; lymphoma; carcinoma; metastasis; ss.

XX OS Homo sapiens.

XX PN WO200009525-A2.

XX PD 24-FEB-2000.

XX PF 03-AUG-1999; 99WO-US171712.

XX PR 03-AUG-1998; 98US-0095212.

XX PA (UYEC-) UNIV EAST CAROLINA.

XX PI Nyce JW;

XX DR WPI; 2000-205971/18.

XX PT New antisense oligonucleotides useful for treating e.g. pulmonary vasoconstriction, inflammation, allergies, asthma, hypertension, bronchitis, emphysema, respiratory distress syndrome, ischemia or cancers -

XX PS Disclosure; Page 900-903; 1343pp; English.

CC The present invention describes a new composition comprising an antisense oligonucleotide (ON) with low adenosine (up to 15%), which targets nucleic acids involved in bronchoconstriction, allergies, and/or inflammation. The ON can have antiinflammatory, antiallergic, antiasthmatic, cytosstatic and analgesic activities. The compositions are useful for the treatment of diseases associated with inflammation, impaired airways, including lung disease and diseases whose secondary effects afflict the lungs of a subject. They can be used for treating e.g. ischaemic conditions, pulmonary vasoconstriction, allergies, asthma, impaired respiration, respiratory distress syndrome, pain, cystic fibrosis, pulmonary hypertension, emphysema, chronic obstructive pulmonary disease (COPD), and cancers such as leukaemias, lymphomas, carcinomas, and cancers which may metastasise to the lungs, including breast and prostate cancer. The reduction of the adenosine content of the ONs reduces side effects. The A-containing ONs break down with the release of deoxyadenosine which activates adenosine receptors causing bronchoconstriction and inflammation. AAA32313 to AAA35312 represent the nucleotide sequences given in the sequence listing from the present invention, which correspond to SEQ ID NO:1 to 2815, and then the last 185 sequences are also called SEQ ID NO:1 to 185, but the sequences differ from the previously named sequences. SEQ ID NO:11 to 1680 (AAA32323 to AAA33992) are specifically claimed ONs from the present invention. N.B. Sequences given in the disclosure of the present invention do not match up with their corresponding SEQ ID NO: sequences given in the sequence listing.

XX SQ Sequence 10032 BP; 2342 A; 2486 C; 2643 G; 2561 T; 0 other;

Query Match 7.0%; Score 185.8; DB 21; Length 10032;
Best Local Similarity 55.7%; Pred. No. 5.1e-28;
Matches 443; Conservative 0; Mismatches 337; Indels 15; Gaps 4;

QY 1869 aaaaattagctggcggtggatgcatccacaatcccagctactgggagggctgagcc 1928
DB 8514 AAAAAATTAGCGGGCATGGTGGCATGTATAGTCCCAGCTACTCGGAGGCTGAGGC 8455
QY 1929 atgagaatcgttgaaccggggagcgagatgttgcagtgaagccgagacgagccactgca 1988
DB 8454 AGGAGATGGCGTGAACCTGGGAGCGGAGCTTGCAGTGAGCGGCAGATCGGCCACTGCA 8395
QY 1989 ctccagctggactacagagcgagactctctcaaaaa-aaaaaataaaataaaagtaa 2047
DB 8394 CTCAGCTGGGCGACAGAGCGAACTCCATCTCAAAAAGAGTTACACATATAGAATGA 8335
QY 2048 cttaggtgcaagggtgctctctgttattcaactgaagaccgtgccccgggttatgaggtgtac 2107
DB 8334 ATTGCGCCCTCTCCCAAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAA 8275
QY 2108 cagaagcaagatttcaactatgcacactattcacgcgtccaccctagcattgaagccagcc 2167
DB 8274 ACCCAACCAACCGCATGTTATTATAGTAAACAGCAGCTTAAATGGGTGGGCAC 8215
QY 2168 ttagcctgaaagccttggcttggagggcgaggtcttcccccaaaatgcagacacgaaggt 2227
DB 8214 AGTGGCTCATGCTGTAATCCAGCACTTCGGGGGGTCAAGGCAGCGAGATCATCTTGAG 8155
QY 2228 gcaagtgaagctgcagctcttgcaaaaagatgtaacttgtc-----acgaagggccac 2279
DB 8154 CCCAGAGATTTCAGAGACCAGCGCTGGCAACATGGTAAATCCCGCTCTACTAATAATACACA 8095
QY 2280 gagtggcagggagagctgtccacatcttgcggaagtgtgctatgtgagacggggagggcg 2339
DB 8094 ATTTAGCTGGGAGTGGTGGTGCATGCTGCTGTAGTTCAGCTACTTTGAG--AGGCTGAGGCA 8037
QY 2340 ggtcccttagagataagagacaatacataaaggggagagatatcagagaaaaatcgtaaaggag 2399
DB 8036 GAAGAATTGCTTGAACCCGGGAGGAGGTTCAGTGAGTGAGATGGCACCACTGCACATCC 7977
QY 2400 cagatggttgcagagagaataggtgacctcgaagagactggcgagagcttctcagaaaaac 2459
DB 7976 AACCTGGGAGATAAGAG---CGAGATCCAGTCTCAGAAAAAGAAAAAAGG 7921
QY 2460 cactggagcgtggcgacagctgagccttaggcctgtaatcccagcacttgggaggtgacg 2519
DB 7920 ACTTAATTGGTTGGCGGAGTGGCTCATGCTGAATCCAGCACTTTGGGGGGCCAAAGG 7861
QY 2520 cagdtgaatcacttgaggtcaggttcagaccagcctggcgcaacatggtgaaccccca 2579
DB 7860 CAGGTGGTTCACCTTAAGTCAAGAGTTCAAGACCAAGCTTGGCCAAACATGATGAACCCCG 7801
QY 2580 tctctacagaaaaatataaaaaattagcagggcggtggtggcacaagcctagaatcccagcta 2639
DB 7800 TCTCCACAAAAATACAAAAAATTAGCCAGGCATGCTGGCACACATCTGTGTCTCAGCTA 7741
QY 2640 ctgggaggtgaggg 2654
DB 7740 CTCAGGAGGCCAAGG 7726

Search completed: January 11, 2002, 17:19:41
Job time: 29533 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 11, 2002, 11:36:18 ; Search time 8779.43 Seconds
(without alignments)
3248.422 Million cell updates/sec

Title: US-09-820-005-3_COPY_16200_18853
Perfect score: 2654
Sequence: 1 tgcctagccatagctggc.....agctacttgaggagctgagg 2654

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:
1: em_estfun:*
2: em_esthum:*
3: em_estin:*
4: em_estom:*
5: em_estpl:*
6: em_estba:*
7: em_estro:*
8: em_estov:*
9: em_hic:*
10: gb_est1:*
11: gb_est2:*
12: gb_hic:*
13: gb_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_inv:*
17: em_gss_pln:*
18: em_gss_pro:*
19: em_gss_rod:*
20: em_gss_vrt:*
21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	420.4	15.8	538	13	AQ223799 HS_2218_A
C 2	335.6	12.6	476	10	A1760998 wh97c08.x
C 3	332.6	12.5	478	10	A1540062 td08g08.x
C 4	332.6	12.5	594	10	AW515919 xy02c11.x
C 5	331.2	12.5	444	10	AW015376 UI-H-B10-
C 6	331.2	12.5	750	10	AW072388 xa07d05.x
C 7	330.4	12.4	441	10	A1239613 qh37d07.x
C 8	330	12.4	561	10	A1682248 wa71g04.x
C 9	329.4	12.4	451	10	AW572661 xx92b07.x
C 10	329.4	12.4	625	10	A1479089 tm31a08.x
C 11	318.2	12.0	337	10	AW977599 EST389708
C 12	287.4	10.8	311	10	AW207763 UI-H-B11-

C 13	281.4	10.6	547	10	A1436579
C 14	280.8	10.6	877	10	AA767386 nz81hl1.s
C 15	280.6	10.6	435	10	AI076022 oy60a07.x
C 16	280.6	10.6	464	10	AI086029 qy70g08.x
C 17	280.6	10.6	530	10	AI239858 qh33a01.x
C 18	280.6	10.6	535	10	AI439666 tc91h03.x
C 19	280.2	10.6	356	10	AI933814 wn92f07.x
C 20	280	10.6	771	10	AI990864 ws32g06.x
C 21	279.6	10.5	467	10	AI356863 qy18a06.x
C 22	279.6	10.5	535	11	BE855380 7g13a05.x
C 23	279	10.5	398	10	AI084212 oy72e10.x
C 24	279	10.5	418	10	AI657120 tt49g11.x
C 25	279	10.5	454	10	AW520292 hf38f01.x
C 26	279	10.5	465	10	AI378090 tc80a08.x
C 27	279	10.5	477	10	AI804631 tc81g07.x
C 28	279	10.5	485	10	AI364917 qz23a06.x
C 29	279	10.5	488	10	AI802729 wf17f06.x
C 30	279	10.5	551	10	BE501539 bw33e06.x
C 31	278.4	10.5	466	11	BF438587 nab94g01.
C 32	277.4	10.5	323	10	AI633917 tt23a09.x
C 33	277.4	10.5	490	11	BF510804 UI-H-B14-
C 34	277	10.4	292	13	AF166290 AF166290
C 35	276.8	10.4	566	11	BF064049 7h93b04.x
C 36	276.4	10.4	563	10	AI161212 qb55b11.x
C 37	276.2	10.4	508	10	AI524254 th11f03.x
C 38	275.8	10.4	328	10	AI633353 ts96b09.x
C 39	275.8	10.4	467	10	AI356259 qz25e02.x
C 40	275.8	10.4	478	10	BE674817 7e95b08.x
C 41	275.8	10.4	497	10	BE675439 7f09b05.x
C 42	275	10.4	594	10	AI914983 wf81e12.x
C 43	274.6	10.3	507	10	AI218309 qh16b07.x
C 44	274.2	10.3	461	10	BE646379 7e85h01.x
C 45	274.2	10.3	547	10	AI682378 wc52e03.x

ALIGNMENTS

RESULT 1
LOCUS AQ223799/c 538 bp DNA GSS 20-SEP-1998
DEFINITION HS_2218_AL_C06_MR CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate-2218 Col-11 Row=E, DNA sequence.
ACCESSION AQ223799
VERSION AQ223799.1 GI:3639642
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 538)
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., and Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.
TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE 99380589
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallaceu.washington.edu
Sequence Tagged Connector
Plate: 2218 row: E column: 11
Class: BAC ends
High quality sequence stop: 538.
LOCATION/Qualifiers
1..538
/organism="Homo sapiens"
/db_xref="taxon:9606"

```
/clones="Plate=2218 Col=11 Row=E"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/notes="Organ: sperm; Vector: pBelBAC11; BAC Clones in
E-Coli DH10B"
BASE COUNT      116 a   144 c   117 g   149 t   12 others
ORIGIN

Query Match      15.8%; Score 420.4; DB 13; Length 538;
Best Local Similarity 94.8%; Pred. No. 4.3e-46;
Matches 455; Conservative 0; Mismatches 20; Indels 5; Gaps 2;

Qy 1971 cgaagcggccactgactccactccgctggactacagagcagactctatctcaaaaaa 2030
Db 520 CCAGACGGCGCA-TGACTCCACTGGACTAGAGCGAGACTCATCTCAAAAAA 462
Qy 2031 aaaaaaataaactaggtgcagggtgctctgtttattcaatgagaccgtgcc 2090
Db 461 NAANAAAAA---GTACTTAGGTGCAGGGTCTCTCTGTTATTCACTGAGACCGTGCC 406
Qy 2091 cggttatgaggttacagaaacaagattatcaactatgacactattcaaccgtcacc 2150
Db 405 CGGTTATGAGGTTTACAGAAAGCAAGTATTCATATGACACTATTCACCGCTCACT 346
Qy 2151 tagcattgaagccgctgtgactgaaagcctttgtttgaggcaggtctttccccaa 2210
Db 345 TAGCATTGAAGCCAGCTGTAGCTGTANAGCTTTGCTTTGAGGCGAGGCTTTCCCAA 286
Qy 2211 aatcagacacgaagggtgcaagtgaaagctgcagctgttgcgaagaagtgtatgtcac 2270
Db 285 AATCGACACAGAAAGGTGCAAGTGAAGCTGCCAGTCTTGCAAAATATGTAACTTGTAC 226
Qy 2271 gaagccacagtgaggcagagagactgtccacattgcgaagtgcattatgaagacg 2330
Db 225 GAAGCCCCGAGCGGAGTAGACTGTCCACATTTTCGGAAGTGCTATGTGAGGACG 166
Qy 2331 ggggagggggtcccttagagataagagacaatacataaggggagataatcagagaaaaatcg 2390
Db 165 GGGGAGGCGGCTCCCTTAGAGATAAGATACAATCAATGAAGGGAGATATCAGAGAAATCG 106
Qy 2391 taaggagagagatggtgttcaagaaataggtgacatcagaagcgtggcagaagctt 2450
Db 105 TAAGGAACACAGATGGTGTGTCAGAGAAATAGGCTGACCATCGAAGTACTGCGCAGGAGC 46

RESULT 2
A1760998/c
LOCUS
DEFINITION
wh97c08.x1 NCI-CGAP_CLL1 Homo sapiens cdna clone IMAGE:2388686 3',
similar to gb:M55067 NEUTROPHIL CYTOSOL FACTOR 1 (HUMAN);, mRNA
sequence.
ACCESSION
A1760998
VERSION
A1760998.1 GI:5176665
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 476)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgaps-r@mail.nih.gov
Tissue Procurement: Ash Alizadeh, John Byrd, M.D., Mike Grever,
M.D., Louis M. Staudt, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
```

```
www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 1388 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 388.
FEATURES
Location/Qualifiers
source
1..476
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2388686"
/tissue_type="B-cell, chronic lymphocytic leukemia"
/lab_host="DH10B"
/notes="vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I oligo(dT) primer [5',
TGTACCAATCTGAAGTGGAGCGGCGCATGTCTTTTTTTTTTTTTTTTTTTTTT
T 3']; double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified pT73 vector.
Library is normalized, and was constructed by Bento
Soares and M.Fatima Bonaldo."
BASE COUNT      70 a   152 c   166 g   88 t
ORIGIN

Query Match      12.6%; Score 335.6; DB 10; Length 476;
Best Local Similarity 89.2%; Pred. No. 5.1e-35;
Matches 362; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

Qy 823 cagagagcgtgtggcgggccagtggtgcggggcggtctgactcgcccgctct 882
Db 412 CAGCGTCCGTTTCTGCAGCAGCAGCCGCCAGCGCGGGGACCGCAGAGCCCCGGG 353
Qy 883 ctgcccgcagagagagcggcagcagcagcgtctaaacgcagcggcggtcccccg 942
Db 352 AGCCCGCTCGAGGAGGAGCGGCGCAGCAGCGCTTTAAACCGCAGCGCGGTGCCCCCG 293
Qy 943 cggccgagcgcgcacctcatctctgaaacgctgcagcagagacacacgaagcgagcgtg 1002
Db 292 CGGCGGAGGCGCGACCTCATCTCAACCGCTGCAGCGAGAGCACCAGCGAGCTGGCG 233
Qy 1003 tctgcgctgagagcgtggagcgcagtcgccagctagcgtctcgcccttgcgcggcggtg 1062
Db 232 TCTGCCGTCTGAGGCTGAGCGCAGTCCCGAGCTAGGCTCTCGGCCCTTGCGCGCGCGTG 173
Qy 1063 cctgtatatcgttcttatagacctgagctctggacgcgcagcggcagcggcagccct 1122
Db 172 CCTGTATATAGTGTCTATAGAGCCTGGCGCTCTGGAGCGCGGAGGCGAGCCCGACCCCT 113
Qy 1123 gtcagcgcggtcccccaccctcaataaagtgtgcttgagtgagcggaggtctgca 1182
Db 112 GTCCAGCGCGGCTCCCGCCACCCCTCAATAAATGTTGCTTGAGTGGACCGAGGCTCTGCA 53
Qy 1183 ggaatgcaggagggcggtcgcccgcccgaggggtatttttaagt 1228
Db 52 GGAATGCAGGAGGCGCGGCTCGCCCGCCAGGGTTATTTTAAATT 7

RESULT 3
A1540062/c
LOCUS
DEFINITION
td08g08.x1 NCI-CGAP_CLL1 Homo sapiens cdna clone IMAGE:2075102 3',
similar to gb:M55067 NEUTROPHIL CYTOSOL FACTOR 1 (HUMAN);, mRNA
sequence.
ACCESSION
A1540062
VERSION
A1540062.1 GI:4457435
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 478)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
```

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps-r@mail.nih.gov
Tissue Procurement: Ash Alizadeh, John Byrd, M.D., Mike Grever,
M.D., Louis M. Staudt, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert length: 1586 Std Error: 0.00
Seq primer: -400P from Gibco
High quality sequence stop: 378.

FEATURES
source
1..478
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2075102"
/clone_lib="NCI-CGAP_CELL1"
/tissue_type="B-cell, chronic lymphocytic leukemia"
/lab_host="DH10B"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5',
TGTTACCAATCTGAAGTGGAGCGCGCATCTCTTTTCTTTTCTTTTCTTTT
T 3']; double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified pT73 vector.
Library is normalized, and was constructed by Bento
Soares and M.Fatima Bonaldo."
BASE COUNT 69 a 153 c 174 g 82 t

Query Match 12.5%; Score 332.6; DB 10; Length 478;
Best Local Similarity 93.5%; Pred. No. 1.3e-34;
Matches 347; Conservative 0; Mismatches 24; Indels 0; Gaps 0;
Qy 851 gcggggcgggcgctgactgagcccgctctctgcccagagagagcgagcgagcgc 910
Db 375 gccagggcgggcgagcccgagagcccgagagcccgagagcgagcgagcgc 316
Qy 911 agcgctctaaaccgagcgagcgagcgagcgagcgagcgagcgagcgagcgc 970
Db 315 agcgctctaaaccgagcgagcgagcgagcgagcgagcgagcgagcgagcgc 256
Qy 971 gctgagcgagagcaccagcgagcgagcgagcgagcgagcgagcgagcgagcgc 1030
Db 255 gctgagcgagagcaccagcgagcgagcgagcgagcgagcgagcgagcgagcgc 196
Qy 1031 ccagctagcgctgctgagcgagcgagcgagcgagcgagcgagcgagcgagcgc 1090
Db 195 ccagctagcgctgctgagcgagcgagcgagcgagcgagcgagcgagcgagcgc 136
Qy 1091 gcgctgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgc 1150
Db 135 cggtctggagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgc 76
Qy 1151 aaatgttgccttgagtgagcgagcgagcgagcgagcgagcgagcgagcgagcgc 1210
Db 75 aaatgttgccttgagtgagcgagcgagcgagcgagcgagcgagcgagcgagcgc 16

Qy 1211 cagggttattt 1221
Db 15 cagggttattt 5
RESULT 4
AWS15919/c 594 bp mrna EST 03-MAR-2000
LOCUS AWS15919

DEFINITION xy02cll.x1 NCI-CGAP_Lym12 Homo sapiens cDNA clone IMAGE:2851988 3'
similar to gb:M55067 NEUTROPHIL CYTOSOL FACTOR 1 (HUMAN);, mrna
sequence.
ACCESSION AW515919
VERSION AW515919.1 GI:7154001
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 594)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps-r@mail.nih.gov
Life Technologies catalog #: 11547-015
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
image.llnl.gov/image/html/iresources.shtml
Seq primer: -400P from Gibco
High quality sequence stop: 425.

FEATURES
source
1..594
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2851988"
/clone_lib="NCI-CGAP_Lym12"
/tissue_type="lymphoma, follicular mixed small and large
cell"
/lab_host="DH10B"
/note="Organ: lymph node; Vector: pCMV-SPORT6; Site_1:
SalI; Site_2: NotI; Cloned unidirectionally. Primer:
Oligo dT. Average insert size 1.25 Kb. Life Technologies
catalog #: 11547-015"
BASE COUNT 85 a 187 c 209 g 113 t

Query Match 12.5%; Score 332.6; DB 10; Length 594;
Best Local Similarity 93.5%; Pred. No. 1.2e-34;
Matches 347; Conservative 0; Mismatches 24; Indels 0; Gaps 0;
Qy 851 gcggggcgggcgctgactgagcccgctctctgcccagagagagcgagcgagcgc 910
Db 381 gccagggcgggcgagcccgagagcccgagagcccgagagcgagcgagcgc 322
Qy 911 agcgctctaaaccgagcgagcgagcgagcgagcgagcgagcgagcgagcgc 970
Db 321 agcgctctaaaccgagcgagcgagcgagcgagcgagcgagcgagcgagcgc 262
Qy 971 gctgagcgagagcaccagcgagcgagcgagcgagcgagcgagcgagcgagcgc 1030
Db 261 gctgagcgagagcaccagcgagcgagcgagcgagcgagcgagcgagcgagcgc 202
Qy 1031 ccagctagcgctgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgc 1090
Db 201 ccagctagcgctgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgc 142
Qy 1091 gcgctgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgc 1150
Db 141 gcgctgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgc 82
Qy 1151 aaatgttgccttgagtgagcgagcgagcgagcgagcgagcgagcgagcgagcgc 1210
Db 81 aaatgttgccttgagtgagcgagcgagcgagcgagcgagcgagcgagcgagcgc 22
Qy 1211 cagggttattt 1221
Db 21 cagggttattt 11

```

RESULT 5
AW015376/c
LOCUS
DEFINITION
  UI-H-BIO-aat-c-12-0-UI.s1 NCI_CGAP_Sub1 Homo sapiens cDNA clone
  IMAGE:2710295 3', mRNA sequence.
ACCESSION
  AW015376
VERSION
  AW015376.1 GI:5864133
KEYWORDS
  EST.
SOURCE
  human.
ORGANISM
  Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
  1 (bases 1 to 444)
  NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
  Tumor Gene Index
  Unpublished (1997)
  Contact: Robert Strausberg, Ph.D.
  Email: cgaps-r@mail.nih.gov
  The sequence contained an oligo-dT track that was present in the
  oligonucleotide that was used to prime the synthesis of first
  strand cDNA and therefore this may represent a bonafide poly A
  tail. cDNA Library Preparation: M.B. Soares lab Clone distribution:
  NCI-CGAP clone distribution information can be found through the
  I.M.A.G.E. Consortium/LLNL at:
  www-bio.llnl.gov/bbrp/image/image.html
  Seq primer: M13 Forward
  POLYA=Yes.
FEATURES
  Location/Qualifiers
    1..444
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /clone="IMAGE:2710295"
      /clone_lib="NCI_CGAP_Sub1"
      /lab_host="DH10B (Life Technologies)"
      /note="vector: pT73D-Pac (Pharmacia) with a modified
      polylinker; Site.1: Not I; Site.2: Eco RI; The
      NCI_CGAP_Sub1 library is a subtracted library derived from
      BI. BI constitutes a mixture of 21 normalized or
      subtracted NCI_CGAP libraries: NCI_CGAP_Co4,
      NCI_CGAP_Pr22, NCI_CGAP_Pr28, NCI_CGAP_Co10, NCI_CGAP_Co16
      , NCI_CGAP_Kid15, NCI_CGAP_Kid12, NCI_CGAP_Kid3,
      NCI_CGAP_Kid11, NCI_CGAP_Lym2, NCI_CGAP_Br2, NCI_CGAP_Co8,
      NCI_CGAP_CLL1, NCI_CGAP_Le12, NCI_CGAP_Brn23, NCI_CGAP_Lu5
      , NCI_CGAP_Lu24, NCI_CGAP_Lu19, NCI_CGAP_GC4, NCI_CGAP_GC6
      , NCI_CGAP_Brn25. These 21 libraries were pooled and a
      single-stranded DNA preparation of the resulting mixture
      was used as a tracer in a subtractive hybridization with a
      driver whose composition is detailed below: NCI_CGAP_Kid3
      pool 1 LLAM 3334-3337, 3682-3683, 3798-3803 (IMAGE
      Clonoids 132376-132391, 1456008-1456775, 1500552-1502855
      ) NCI_CGAP_Kid5 pool 1 LLAM 3338-3342, 3722-3725,
      3776-3778 (IMAGE Clonoids 1323912-1325831,
      1471368-1472903, 1492104-1493255) NCI_CGAP_Lu5 pool 1 LLAM
      3575-3582, 3851-3854 (IMAGE Clonoids 1414920-1417991,
      1520904-1522439) NCI_CGAP_GC4 pool 1 LLAM 3164-3167,
      3716-3720, 3733-3735 (IMAGE Clonoids 1257096-1258631,
      1469064-1470983, 1475592-1476743) NCI_CGAP_Pr22 pool 1
      LLAM 2457-2459, 2758-2759, 3062-3068 (IMAGE Clonoids
      983608-986759, 1101192-1101959, 1217928-1220615)
      NCI_CGAP_Co10 pool 1 LLAM 2644-2653, 2871-2872 (IMAGE
      Clonoids 1057416-1061255, 1144584-1145351) The resulting
      subtracted library contained 530,000 recombinants.
      Subtraction was performed as previously described [Bonaldi
      , Lennon & Soares (1996): Normalization and Subtraction:
      Two Approaches To Facilitate Gene Discovery. Genome
      Research 6, 791-806.
      TAG LIB=NCI_CGAP_Br2
      TAG TISSUE=breast
      TAG_SEQ=RAACC"
    62 a 140 c 155 g 87 t

```

```

Query Match 12.5%; Score 331.2; DB 10; Length 444;
Best Local Similarity 92.6%; Pred. No. 2e-34;
Matches 348; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

QY 851 gcggggcgggcggtgactcgcccgctctctgcccgcagagagagcgagcgcgc 910
  ||||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 390 GCCAGGCGCGCGGACCGAGAGCCCGGAGAGCCCGCTCGAGGAGGAGGCGAGAGCC 331
  ||||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 911 agcgtctaaaccgcagccggtgccccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 970
  ||||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 330 AGCGCTCTAAACCGAGCGCGGTGCCCCCGCGCGCGCGCGCGCGCGCGCGCGCGCG 271
  ||||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 971 gctcagcagagcaccacgaagcgtggcgctgcgctgaggtgaggtgagcgcagtc 1030
  ||||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 270 GCTGAGGAGGAGGACCAAGCGGAGCTGGCGTCTGAGGTGGAGCGCGAGTCC 211
  ||||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1031 ccagctagctgtcgcccttgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1090
  ||||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 210 CCAGCTAGCGTCTCGGCCCTTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 151
  ||||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1091 gcctgtgacccgagggcagccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1150
  ||||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 150 GCGTCTGGACCGCGAGGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 91
  ||||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1151 aaatgttgcgttgagtgacccagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1210
  ||||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 90 AAATGTTGCTTGGAGTGACCGCGAGGCTCTGCAGGAATGCAGGAGGCGCGCGCGCGCGCG 31
  ||||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1211 cagggttattttcttaa 1226
  ||||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 30 CAGGGTATTTCTTAA 15
  ||||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | |

```

```

RESULT 6
AW072388/c
LOCUS
DEFINITION
  xa07405.x1 Soares_NFL_T_GBC.S1 Homo sapiens cDNA clone
  IMAGE:2567625 3' similar to gb:M55067 NEUTROPHIL CYTOSOL FACTOR 1
  (HUMAN); mRNA sequence.
ACCESSION
  AW072388
VERSION
  AW072388.1 GI:6027386
KEYWORDS
  EST.
SOURCE
  human.
ORGANISM
  Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
  1 (bases 1 to 750)
  NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
  Tumor Gene Index
  Unpublished (1997)
  Contact: Robert Strausberg, Ph.D.
  Email: cgaps-r@mail.nih.gov
  This clone is available royalty-free through LLNL; contact the
  IMAGE Consortium (info@image.llnl.gov) for further information.
  Insert Length: 1386 Std Error: 0.00
  Seq primer: -400P from Gibco
  High quality sequence stop: 464.
  Location/Qualifiers
    1..750
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /clone="IMAGE:2567625"
      /clone_lib="Soares_NFL_T_GBC.S1"
      /lab_host="DH10B"
      /note="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with
      a modified polylinker; Site.1: Not I; Site.2: Eco RI;
      Equal amounts of plasmid DNA from three normalized
      libraries (fetal lung NBHL19W, testis NHT, and B-cell
      NCI_CGAP_GCBI) were mixed, and ss circles were made in
      vitro. Following HAP purification, this DNA was used as

```

BASE COUNT
ORIGIN

tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 3 libraries. The pools consisted of I.M.A.G.E. clones 297480-302087, 682632-687239, 726408-728711, and 729096-731399. Subtraction by Bento Soares and M. Fatima Bonaldo. "

BASE COUNT 114 a 233 c 254 g 148 t 1 others

ORIGIN

Query Match 12.5%; Score 331.2; DB 10; Length 750;
Best Local Similarity 92.6%; Pred. No. 1.7e-34;
Matches 348; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

QY 851 qcggggcgggcggtctgactcgcccgctctctgcccgcagagagagagcgagcgc 910
|||||
Db 377 GCCAGGCGCGCGGGACCGAGAGCCCCGGGAGCCCGCTCGAGGAGGCGCGAGACGC 318
QY 911 agcgtctaaacgcagcgccggtgccccgcggcgagcgccgcacctcatctgaacc 970
|||||
Db 317 AGCGCTTTAAACCGCAGCGCGGTGCCCGCGCGGAGCGCGACCTCATCTGAACC 258
QY 971 gctgcagcgagagcaccagcgaagctgcgctctgcccgtgaggtgagcgagctcc 1030
Db 257 GCTGCAGCGAGAGCACCAGCGGAAGCTGGCGTCTGCCGCTCGAGGCTGGAGCGCAGTCC 198
QY 1031 ccaactagctctgccccttgcggcccgctgctgtatagctgtctatagagcctg 1090
|||||
Db 197 CCAGCTAGCGTCTCGGCCCTTGCCCGCGCGCTGTATATACGTGTCTATAGAGCGCTG 138
QY 1091 gcgtctgagcgagcgagcgcccgaccctgtccagcgcggtcccgccaccctcaat 1150
|||||
Db 137 GCGTCTGGACGCGGAGGCGAGCCCGACCCCTGTCCAGCGCGGCTCCCGCACCTCAAT 78
QY 1151 aaatgttcttgagtgagcagcgagctctcaggaatgcaggagggcggtcccgccc 1210
Db 77 AAATGTTGCTTGGAGTGACCGAGGCTCTGCAGGAATGCAGGGAGGCGCGGCTCCGCC 18
QY 1211 cagggttatttctaa 1226
|||||
Db 17 CAGGGTTATTCTTAA 2

RESULT 7
AI239613/c
LOCUS AI239613 441 bp mRNA EST 13-APR-1999
DEFINITION qb37407.x1 Soares_NFL_T_GBC.S1 Homo sapiens cDNA clone
IMAGE:1846861.3' similar to gb:M55067 NEUTROPHIL CYTOSOL FACTOR 1
(HUMAN);, mRNA sequence.
ACCESSION AI239613
VERSION AI239613.1 GI:3835010
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 441)
AUTHORS NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapb-r@mail.nih.gov
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 1489 Std Error: 0.00
Seq primer: -400p from Gibco
High quality sequence stop: 324.
FEATURES Location/Qualifiers
source 1..441
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1846861"

/clone_lib="Soares_NFL_T_GBC.S1"
/lab_host="DH10B"
/notes="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with
a modified polylinker; Site 1: Not 1; Site 2: Eco RI;
Equal amounts of plasmid DNA from three normalized
libraries (fetal lung NBHL19W, testis NHR, and B-cell
NCI-CCAP-GCBI) were mixed, and ss circles were made in
vitro. Following HAP purification, this DNA was used as
tracer in a subtractive hybridization reaction. The driver
was PCR-amplified cDNAs from pools of 5,000 clones made
from the same 3 libraries. The pools consisted of
I.M.A.G.E. clones 297480-302087, 682632-687239,
726408-728711, and 729096-731399. Subtraction by Bento
Soares and M. Fatima Bonaldo. "

BASE COUNT 65 a 145 c 152 g 79 t

ORIGIN

Query Match 12.4%; Score 330.4; DB 10; Length 441;
Best Local Similarity 93.0%; Pred. No. 2.5e-34;
Matches 346; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

QY 851 qcggggcgggcggtctgactcgcccgctctgcccgcagagagagagcgagcgc 910
|||||
Db 383 GCCAGGCGCGCGGACCGCAGAGCCCCGGGAGCCCGCTCGAGGAGGCGCGAGACGC 324
QY 911 agcgtctaaacgcagcgcggtgccccgcggcgagcgccctcatctgaacc 970
|||||
Db 323 AGCGCTTTAAACCGCAGCGCGGTGCCCGCGCGGCGCGACCTCATCTGAACC 264
QY 971 gctgcagcgagagcaccagcgaagctggcgctctgaggtggagcgagctcc 1030
|||||
Db 263 GCTGCAGCGAGAGCACCAGCGGAAGCTGCGCTGTGCCGCTCGAGGCTGGAGCGCAGTCC 204
QY 1031 ccaactagctctgccccttgcggcccgctgctgtatagctgtctatagagcctg 1090
Db 203 CCAGCTAGCGTCTCGGCCCTTGCCCGCGCGCTGTATATACGTGTCTATAGAGCGCTG 144
QY 1091 gcgtctgagcgagcgagcgcccgaccctgtccagcgcggtcccgccaccctcaat 1150
|||||
Db 143 GCGTCTGGACGCGGAGGCGAGCCCGACCCCTGTCCAGCGCGGCTCCCGCACCTCAAT 84
QY 1151 aaatgttcttgagtgagcagcgaggtctctgcaggaatgcaggagggcggtcccgccc 1210
Db 83 AAATGTTGCTTGGAGTGACCGAGGCTCTGCAGGAATGCAGGGAGGCGCGGCTCCGCC 24
QY 1211 cagggttatttt 1222
|||||
Db 23 CAGGGTTATTATT 12

RESULT 8
AI682248/c
LOCUS AI682248 561 bp mRNA EST 07-MAR-2000
DEFINITION wa71904.x1 Soares_NFL_T_GBC.S1 Homo sapiens cDNA clone
IMAGE:2301654.3' similar to gb:M55067 NEUTROPHIL CYTOSOL FACTOR 1
(HUMAN);, mRNA sequence.
ACCESSION AI682248
VERSION AI682248.1 GI:4892430
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 561)
AUTHORS NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapb-r@mail.nih.gov
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 625)
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapb-r@mail.nih.gov
 Tissue Procurement: Ash Alizadeh, John Byrd, M.D., Mike Grever, M.D., Louis M. Staudt, M.D., Ph.D.
 cDNA Library Preparation: M. Bento Soares, Ph.D.
 cDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www.bio.llnl.gov/bbrp/image/image.html
 Insert Length: 705 Std Error: 0.00
 Seq primer: -40UP from Gibco
 High quality sequence stop: 460.
FEATURES Location/Qualifiers
 source
 1..625
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:2158166"
 /clone_lib="NCI_CGAP_CLL1"
 /tissue_type="B-cell, chronic lymphoc leukemia"
 /lab_host="DH10B"
 /note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 5' strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTCATATCTAAGTGGAGCGCGCATGCTCTTTTTTTTTTTTTTTT 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library is normalized, and was constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT 88 a 198 c 221 g 117 t 1 others
ORIGIN
 Query Match 12.4%; Score 329.4; DB 10; Length 625;
 Best Local Similarity 93.0%; Pred. No. 3e-34;
 Matches 345; Conservative 0; Mismatches 26; Indels 0; Gaps 0;
 Qy 851 qcggggcgggcgctctgaactgcggccgctctctgcgcgcagaggagagcgagcgc 910
 Db 382 GCCAGCGCGCGCGGGACCGACGAGCCCGGGGCCCGCTCGTGGAGGAGCGGAGACGC 323
 Qy 911 agcgctctaaacgcagccgcggcggtgccccgcgcgcgcgcgcgcctcatcctgaacc 970
 Db 322 AGCGCTCTAAACCGACCGCGGCGTGTCCCCGCGCGCGAGCGCGACCTCATCTGAACC 263
 Qy 971 gctgcagcgagagaccacgcggaagctgcgctgtcgctctgagcgctgagcgagctgc 1030
 Db 262 GCTGCAGCGAGACCAAGCGGAAGTGGCGCTGTCGCTGTGAGGCTGGAGCGCAGTCC 203
 Qy 1031 ccagctagctctcgcccttgccgcgccttgctgtatatacgtgtctatagagcctg 1090
 Db 202 CCAGCTAGCTCTCGGCGCTTGGCGGCCCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 143
 Qy 1091 qcgtctgagcgagcgagcgagccccgcacctgtccagcgcgctcccgccacctcaat 1150
 Db 142 CGGTCTGGACCGCGAGCGGACCGCCCGACCCCTGTCCAGCGCGCGCTCCCGCCACCTCAAT 83
 Qy 1151 aaatgttgcttgagtgagaccgagctctgcaggaatgcaggggagggcgcgctccgccc 1210
 Db 82 AATGTGTCTTGAGTGGACCGAGGCTCTGCAGGAATGCAGGAGAGGGCCGGGCTCCGCC 23
 Qy 1211 cagggttattt 1221
 Db 22 CAGGTAATTT 12

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 11, 2002, 17:18:54 ; Search time 14958.6 Seconds
(without alignments)
2926.974 Million cell updates/sec

Title: US-09-820-005-3_COPY_16200_18853
Perfect score: 2654
Sequence: 1 tdtctaggccatagcttggc.....agctacttggggaggtgagg 2654

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb_ba.*

2: gb_htg.*

3: gb_in.*

4: gb_om.*

5: gb_ov.*

6: gb_pat.*

7: gb_ph.*

8: gb_pl.*

9: gb_pr.*

10: gb_ro.*

11: gb_sts.*

12: gb_sy.*

13: gb_un.*

14: gb_vl.*

15: em_ba.*

16: em_fun.*

17: em_hum.*

18: em_in.*

19: em_om.*

20: em_or.*

21: em_ov.*

22: em_pat.*

23: em_ph.*

24: em_pl.*

25: em_ro.*

26: em_sts.*

27: em_sy.*

28: em_un.*

29: em_vi.*

30: em_htgo_hum.*

31: em_htgo_inv.*

32: em_htgo_rod.*

33: em_htg_hum.*

34: em_htg_inv.*

35: em_htg_rod.*

36: em_htg_other.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

8

Result No.	Score	Query Match	Length	DB ID	Description
1	1934.4	72.9	230552	9	AC005098 Homo sapi
2	1930.8	72.8	131359	9	AC004883 Homo sapi
3	1930.8	72.8	275197	9	AC004166 Homo sapi
4	1927.6	72.6	169604	9	AC003884 Homo sapi
5	1735.2	65.4	124526	9	AC005080 Homo sapi
6	1115.8	42.0	17302	9	AF184614 Homo sapi
7	1052.4	39.7	1646	9	HS47P43S03
8	1039.6	39.2	184558	2	AC092405 Papio cyn
9	1028.6	38.8	1619	9	HS47P40S03
10	1027	38.7	1619	9	HS47P41S04
11	1025.4	38.6	1645	9	HS47PHOX03
12	532.8	20.1	149830	2	AC027219 Homo sapi
13	517.6	19.5	149830	2	AC027219 Homo sapi
14	284.4	10.7	57245	2	AC068263 Homo sapi
15	279	10.5	1340	9	AF330627 Homo sapi
16	278	10.5	1339	9	HUMNCF1A
17	278	10.5	1349	9	HUMNADPHO
18	277.4	10.5	1460	9	BC002816 Homo sapi
19	229.6	8.7	175179	9	AC005237 Homo sapi
20	228.4	8.6	170360	2	AC084773 Homo sapi
21	226	8.5	38429	9	AC000035 Homo sapi
22	225	8.5	239566	9	HS4J03147
23	223.8	8.4	100000	9	AP000502 Homo sapi
24	223.8	8.4	109646	9	HS4J03147
25	223.6	8.4	158460	2	AC022826 Homo sapi
26	222.4	8.4	47205	2	AC087652 Homo sapi
27	221.6	8.3	164138	9	AL162386 Human DNA
28	221.2	8.3	135405	9	AC000025 Homo sapi
29	221.2	8.3	149308	9	AC005527 Homo sapi
30	221.2	8.3	318488	9	AC005529 Homo sapi
31	220.2	8.3	78533	9	HS278N12
32	219.8	8.3	100182	2	AL591863 Homo sapi
33	219.8	8.3	172757	2	AL591848 Homo sapi
34	219.6	8.3	134210	9	AC005052 Homo sapi
35	219.6	8.3	176728	2	AC025544 Homo sapi
36	218	8.2	184694	2	HS536018 Homo sapi
37	217.4	8.2	160972	2	AC021417 Homo sapi
38	217.4	8.2	183535	2	AP002754 Homo sapi
39	217.4	8.2	209545	2	AC087848 Homo sapi
40	217.2	8.2	191385	2	AC025283 Homo sapi
41	216.8	8.2	36037	9	AC003966 Homo sapi
42	216.8	8.2	202324	2	AC068720 Homo sapi
43	216	8.1	142494	9	AL137119 Human DNA
44	216	8.1	150789	9	HS21P7
45	216	8.1	179849	2	AC078975 Homo sapi

ALIGNMENTS

RESULT 1

AC005098

LOCUS

DEFINITION

AC005098

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

AC005098 230552 bp DNA 07-OCT-2000
Homo sapiens BAC clone CTA-350L10 from 7q11.2, complete sequence.

AC005098.2 GI:9211526

HTG.

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 230552)

Toward a complete human genome sequence

Genome Res. 8 (11), 1097-1108 (1998)

99063792

2 (bases 1 to 230552)

Threide, J., Abbott, A., Graves, T., Elliott, G. and Markovic, C.

The sequence of Homo sapiens HAC clone CTA-350L10

Unpublished

REFERENCE
AUTHORS
TITLE
JOURNAL

3 (bases 1 to 230552)
Waterston,R.H.
Direct Submission
Submitted (12-JUN-1998) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA

REFERENCE
AUTHORS
TITLE
JOURNAL

4 (bases 1 to 230552)
Waterston,R.H.
Direct Submission
Submitted (15-JUL-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA

REFERENCE
AUTHORS
TITLE
JOURNAL

5 (bases 1 to 230552)
Waterston,R.
Direct Submission
Submitted (07-OCT-2000) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Jul 15, 2000 this sequence version replaced gi:3212893.

COMMENT

Center: Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc>
Contact: sapiens@watson.wustl.edu

Summary Statistics

Center project name: H_RG350L10

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/GB/CHR7>, send <mailto:egreen@nhgri.nih.gov>, or see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

Clone CTA-350L10 is from a release of the human BAC library C1TB-HS-A. The library contains cloned DNA from human sperm. See: Shizuya et al., Proc. Natl. Acad. Sci. USA 89:8794-7 (1992); U-J. Kim et al., Genomics 34:213-8 (1996). The clone is available from Research Genetics, Inc. (<http://www.resgen.com>).

VECTOR: pBelobAC11

Selection: chloramphenicol

NEIGHBORING SEQUENCE INFORMATION:

Actual start of this clone is at base position 1 of CTA-350L10; actual end is at base position 230552 of CTA-350L10.

The sequence CTA-350L10 from base position 222330 to 222775 is a GA rich region. The sequence is not exact but it is believed to be the best representation of this region. The region was sized by PCR from clone DNA at 650 bp. The region corresponds to restriction digest hindiii: band size 7685 in silico and 7756 real.

FEATURES

source

1..230552
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="7"

/map="7q11.2"
/clone="CTA-350L10"
/clone_lib="C1TB-HS-A"
441..725
/rpt_family="Alu"
783..1149
/note="similar to EST BE000775 (NID:g8261008)"
910..1312
/note="similar to EST R22057 (NID:g776838) yh25f08.s1"
1286..1716
/note="similar to EST H72453 (NID:g1044269) ys04f07.r1"
1609..1864
/rpt_family="Alu"
1984..2271
/note="similar to EST R22113 (NID:g776894) yh25f08.r1"
2337..2682
/note="similar to EST AW383168 (NID:g6887827)"
2437..2558
/note="similar to EST N56845 (NID:g1200735) yw83e04.r1"
2532..2601
/rpt_family="MIR"
3054..3367
/rpt_family="Alu"
3240..3391
/note="similar to EST A1566941 (NID:g4525393) tn25b07.x1"
3588..3845
/note="similar to EST H70120 (NID:g1040326) ys01a09.r1"
3599..4017
/note="similar to EST H89808 (NID:g1080238) yu82g03.r1"
3741..3928
/note="similar to EST N56845 (NID:g1200735) yw83e04.r1"
3816..3932
/note="similar to EST AW523936 (NID:g7166321)"
3819..3932
/note="similar to EST AW503416 (NID:g7139965)"
3820..3932
/note="similar to EST AW197058 (NID:g6476288) xb15f09.x1"
3820..3932
/note="similar to EST AW163398 (NID:g6302431) au94g04.y1"
3871..3932
/note="similar to EST A1700706 (NID:g4988606) we39e12.x1"
3876..3932
/note="similar to EST T31784 (NID:g613882)"
3889..4387
/note="similar to EST T79157 (NID:g697666) yd70b07.s1"
4109..4387
/note="similar to EST A1696334 (NID:g4984234) tw59e05.x1"
4502..4812
/rpt_family="Alu"
5041..5087
/rpt_family="Mariner"
5254..5330
/note="similar to EST AA794004 (NID:g2856959) vr37a02.r1"
5265..5316
/note="similar to EST AW503416 (NID:g7139965)"
5265..5330
/note="similar to EST A1700706 (NID:g4988606) we39e12.x1"
5265..5330
/note="similar to EST AW197058 (NID:g6476288) xb15f09.x1"
5265..5330
/note="similar to EST AW523936 (NID:g7166321)"
5265..5330
/note="similar to EST T31784 (NID:g613882)"
5265..5348
/note="similar to EST AW163398 (NID:g6302431) au94g04.y1"
5284..5330
/note="similar to EST T30644 (NID:g612742)"
5305..5330
/note="similar to EST AW993840 (NID:g8254025)"
5750..6041
/rpt_family="Alu"
5964..6127
/note="similar to EST AA078486 (NID:g1837960)"

repeat_region

misc_feature

misc_feature

misc_feature

repeat_region

misc_feature

misc_feature

misc_feature

repeat_region

repeat_region

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

repeat_region

repeat_region

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

repeat_region

misc_feature

[illegible]

Qy 2341 gtcccttagagataagagacaatcataagggagagatatcagagaaaatcgttaaggggagc 2400
|||||
Db 109941 GTCCCTTAGAGAT--GAGACAATCATAGGGGAGATATCAGAGAAAATCGTAAGGGGAGC 109998
Qy 2401 agatgggtgtcaagagaataagctgaccatcgaaggactggaagaagcttttcagaaaaacc 2460
|||||
Db 109999 AGATGGTGTCTCAAGAGATAAGCTGACCATCGAAGGACTGGCAGAAGCTTTTCAGAAAAACC 110058
Qy 2461 actgagcggctgggacagtgcttagcctgttaatccacagcactttgggaggtgacgc 2520
|||||
Db 110059 ACTGGACGGCTGGGCACAGTGGCTTAGGCCCTGTATCCACACACTTTGGGAGGCTGAGCC 110118
Qy 2521 aggtgaatacacttgaggtcaggtatccagaccagcctggcccaacatggtgaaaccccat 2580
|||||
Db 110119 AGGTGAATCACTTGAGGTTCAGGAGTTCACAGACCAGCCTGGCCAAACATGCTCAAAACCCCAT 110178
Qy 2581 ctctacagaaaaataaaaaatgccagggctgtgtggcacaaagcctagaaatcccagctac 2640
Db 110179 CTCTACAGAAAATAAAAAATTAGCCAGGCGTGGTGGCACAAAGCCTAGAAATCCAGAGTAC 110238
Qy 2641 ttggagagctgag 2654
|||||
Db 110239 TTGGAGGCTGAGG 110252

RESULT 3
AC004166 275197 bp DNA PRI 20-JUL-2000
LOCUS Homo sapiens Chromosome 7 BAC Clone 239c10, complete sequence.
DEFINITION AC004166
ACCESSION AC004166
VERSION AC004166.12 GI:8887011
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 275197)
AUTHORS Ren,Q., Burián,D., Huang,E., Meadows,S., Korenberg,J. and Roe,B.A.
TITLE Homo sapiens Chromosome 7 BAC Clone 239c10
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 275197)
AUTHORS Burián,D.M. and Roe,B.A.
TITLE Direct Submission
JOURNAL Submitted (21-FEB-1998) Department of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
REFERENCE 3 (bases 1 to 275197)
AUTHORS Burián,D.M., Ren,Q., Meadows,S., Huang,E., Korenberg,J. and
Roe,B.A.
TITLE Direct Submission
JOURNAL Submitted (01-JUL-2000) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
REFERENCE 4 (bases 1 to 275197)
AUTHORS Ren,Q., Burián,D., Huang,E., Meadows,S., Korenberg,J. and Roe,B.A.
TITLE Direct Submission
JOURNAL Submitted (20-JUL-2000) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
COMMENT On Jul 1, 2000 this sequence version replaced gi:8779478.
FEATURES
source
1. 275197
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="7"
/clone="239c10"

BASE COUNT 70874 a 66648 c 66737 g 70938 t
ORIGIN

Query Match 72.8%; Score 1930.8; DB 9; Length 275197;
Best Local Similarity 76.5%; Pred. No. 0;

Matches 2030; Conservative 0; Mismatches 569; Indels 55; Gaps 4;
Qy 1 tctctagccatagcttgccagtgccgggaggggctctcagcctgagagagggagcag 60
|||||
Db 49590 TGTCTAGGCCATAGCTTGGCAGTGC CGGGGGGGGGCTCTCAGCCTGCAGAGAGGAGCAG 49649
Qy 61 gacctcagcgggggaaagggctggacgcgcctggcccggtgtgtgggctggcagcggggg 120
Db 49650 GACCCCTCAGGGGGGAAAGGGGCTGGACGCGCTGCGCCGCGGTGTGGGGGTGCACAGGGGG 49709
Qy 121 cgaagaaacgcgcagatgccccgggggctttgggagatgggcagctccagggggctccccc 180
Db 49710 CGGAAGAAAGCGCGATGCCCCGGGGGCTTTGGGGATGGGAGTCCAGGGGGGCTCCCGG 49769
Qy 181 gaggggggacagacagaccgaaggtggtgaggggctgtgaaacccgcggcaggtctctgc 240
Db 49770 GAGAGGGGACACAGACCGAAGGCTGTGAGGGGCGTGGAAACCCGCCAGGCTCTGCT 49829
Qy 241 gcagggaaggtctctgtgacggggggcgcgcgcctctgtccccgcgggggtgtgtgc 300
Db 49830 GCAGGGCAAGGTCCTTGTGCTGACGGGGCAGCGCGCTCTTGTCCCGCGGGGTCTGTC 49889
Qy 301 agactaccggccccctactgcccccaacttctcggaccaggggtgcccattgagtc 360
Db 49890 AGACTACCGGCCCCCTACTGCCCCCACTTCTCTGGACACAGGGGTGCCCATCTGAGTCCC 49949
Qy 361 tgggggagggggcgccctcgggctttgacgacgccccctccccgtggggccaggtcgcca 420
Db 49950 TGGGGGACAGGGGGCGCCCTCGGGCTTTTGACGAGCGCCCGTCCCGCTGGGCAGGTCTCCA 50009
Qy 421 tccgcaacgcgcacagcattccaccagcggtcggaagcgctcagccagggagcgcctatc 480
Db 50010 TCCGCAACGCGCACAGCATCCACACGCGTCCGGAAGCGCTCAGCCAGGACGCTATC 50069
Qy 481 gccgcaacagctcgttttctgacgacgacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 540
Db 50070 GCCGCAACAGCGTCCGTTTTCTGACGACGACGCGCGCGCGCGCGCGCGCGCGCAGAG 50129
Qy 541 gcccgggagccgcctcggtgagtgacgagggagggagggagggagggagggagggaggg 600
Db 50130 GCCCGGGAGCCCGCTCGGTGAGTGCAGCGGGAGAGGGCAGGAAGGGCCGTAGAGG 50189
Qy 601 cggagtcagcgggagagcgggggccagagggcagggccagagtagcgggggcgggaccagag 660
Db 50190 CGGAGTCAAGCGGACAGCGGGGCCAGAGGTAGGCCAGAGTAGCGGGCGCGGACAGAG 50249
Qy 661 ggcgaaatcagagggagagggcggggactgagggcgggggcgagagggagggagggaggg 720
Db 50250 GCGGGAATCAGAGGAGAGAGCGGGGACTGGAGCGGGGCTCAGAGAGGAGGACGCGCTAG 50309
Qy 721 gggcgagcgcatccttaagagcgagtcagagggagggcagagggcagagggagggagggc 780
Db 50310 GGGCGGAGCGATCCTTAAGAGCGCGAGTTCAGAGGAGAGGACCAAGCGGGAGGCGAGGC 50369
Qy 781 cagagcgggagcagaggttgagagccgcgcgcggggcgagggccagagagcgctgtggggc 840
Db 50370 CAGAGCGGGAGCAGCAGAGTTGGAGAAACGCGCGGGGAGGCGCAGAGAGCGCTGTGGCG 50429
Qy 841 gggccagtgctgcggggcgggggcgctgtgactggccccgcctctctcggcgagagggag 900
Db 50430 GGGCCAGTGGGGGGGGGGGCTGTGACTCGCGCCCGCTCTCTGCCCGCAGAGAGGAG 50489
Qy 901 cggcagacgcagcgtcttaaacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 960
Db 50490 CGGCAGACCGACCGCTCTAAACCGCAGCGCGGGGTGCCCGCGCGCGCGCGCGCGCGCTC 50549
Qy 961 atcctgaaccgctgcagcagagcagcaccagcggaagctggcgctctgtgcctctgagggctg 1020
Db 50550 ATCCTGAACCGCTGCAGCGAGAGCACCAGCGGAAGTGGCGTCTGCCCTCTGAGGCTGG 50609
Qy 1021 agcgagtcgccagctagcgtctcgccctctgcggcccgctgcctgtatatacgtgttct 1080
Db 50610 AGCGAGTCCCCAGCTAGGCTCTCGGCCCTTGGCCCGCGTGCCTGTACATAGCTGTTCT 50669


```
QY 2101 gttgtaccagaagaagcatttcaactatgacacattattcaccgctcacccttagcattgaa 2160
Db 112814 GTTGTACCAAGAAGCAAGTATTTCACATATGCACATATTACCCGCTCACCTAGCATTTGAA 112873
QY 2161 gccagccttagcctgaagccttgcctttagggcaggctcttccccaaaatgcagaca 2220
Db 112874 GCCAGCCTTAGCCTGAAGACCTTTGCTTTGAGGCGAGGCTCTTCCCCAAAATGCAGACA 112933
QY 2221 cgaagtgcaaaagtgaagctgcagctcttgcaaaagatgtaacttgcacgaagccacg 2280
Db 112934 CGAAGGTGCAAGTGAAGCTGCCAGCTCTGTGCAAAAGATGTACTTGTGCAGGAAGCCACG 112993
QY 2281 agtgcaggagagcgtgtctccacatttgcggaagtgtctatgtgagcggggagcg 2340
Db 112994 ACTGGCAGGAGAGCTGTCCACATTTGCGGAAGTGGCTATGTGAGGACGGGGAGCGG 113053
QY 2341 gtcccttagagataagacatacgaagggagagatatcagagaaaatcgtaaagggagc 2400
Db 113054 GTCCCTTAGAGAT--GAGACAATCATAAAGGGAGATATCAGAGAAAATCGTAAGGGGAGC 113111
QY 2401 agatggtttcagaagataagctgaccatcgaagcactggcagaagcctttcagaaaaacc 2460
Db 113112 AGATGTTGTCAGAGAATAGGCTCACCATCGAAGACTGGCAGAGAGCTTTCAGAAAACC 113171
QY 2461 actgacggctgggcacagtgcttaggcctgtaatccacgacctttggagcgtgcgc 2520
Db 113172 ACTGGACGGCTGGGCACAGCTGCTTAGGCCTCTAATCCACGACCTTGGGAGGCTGACGC 113231
QY 2521 agtgaataccttgaggtcagagtgctcagacccagcctgcccacatgggtgaaccccat 2580
Db 113232 GGGTGAATACATTTGAGTTCAGAGTTCAGACACCGCTGGCCAACATGTTGTAAGCCCAT 113291
QY 2581 ctctcacagaaataaaaaattagccagcgtgtggcgacagcctagaatcccagctac 2640
Db 113292 CTCACAGAAAATATAAAATTTAGCCAGCGTGGTGGCCAAAGCCTAGAACCCAGCTAC 113351
QY 2641 ttgggagcgtgagg 2654
Db 113352 TTGGGAGGCTGAGG 113365

RESULT 5
LOCUS AC005080 124526 bp DNA PRI 02-OCT-2000
DEFINITION Homo sapiens BAC clone CTA-269p13 from 7q11.2, complete sequence.
ACCESSION AC005080
VERSION AC005080.2 GI:7770715
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 124526)
Sulston,J.E. and Waterston,R.
JOURNAL
Toward a complete human genome sequence
Genome Res. 8 (11), 1037-1108 (1998)
MEDLINE 99063792
REFERENCE 2 (bases 1 to 124526)
AUTHORS Scott,K., Layman,D., Kalicki,J. and Harmon,G.
TITLE The sequence of Homo sapiens BAC clone CTA-269p13
JOURNAL Unpublished
AUTHORS 3 (bases 1 to 124526)
Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (12-JUN-1998) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
4 (bases 1 to 124526)
Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (12-MAY-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
```

REFERENCE
AUTHORS
TITLE
JOURNAL

5 (bases 1 to 124526)
Waterston,R.H.
Direct Submission
Submitted (12-JUN-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
6 (bases 1 to 124526)
Waterston,R.
Direct Submission

REFERENCE
AUTHORS
TITLE
JOURNAL

Submitted (02-OCT-2000) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On May 12, 2000 this sequence version replaced gi:3212911.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: sapiens@wustl.wustl.edu
----- Summary Statistics

Center project name: H_RG269P13

NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.

MAPPING INFORMATION:

The sequence of this clone was established as part of a mapping and
sequencing collaboration between the NHGRI Chromosome 7 Mapping
Project (Eric D. Green, Director), John D. McPherson in the
Department of Genetics (Washington University), and the Washington
University Genome Sequencing Center. For additional information
about the map position of this sequence, see
http://www.nhgri.nih.gov/DIR/OTB/CHR7 , send
mailto:egreen@nhgri.nih.gov , or see http://genome.wustl.edu/gsc

SOURCE INFORMATION:

Clone CTA-269p13 is from a release of the human BAC library
CITB-HS-A. The library contains cloned DNA from human sperm. See:
Shizuya et al., Proc. Natl. Acad. Sci. USA 89:8794-7 (1992); U-J.
Kim et al., Genomics 34:213-8 (1996). The clone is available from
Research Genetics, Inc. (http://www.resgen.com).
VECTOR: pBelOBAC11
Selection: chloramphenicol

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP4-754G14; the clone sequenced 7
to the right is RP11-396K3. Actual start of this clone is at base
position 1 of CTA-269p13; actual end is at base position 124526 of
CTA-269p13.

The clone CTA-269p13 may contain a transposon in the growth of the
clone, which is not part of the submitted sequence.

FEATURES
Source

1. 124526
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="7"
/map="7q11.2"
/clone="CTA-269p13"
/clone.lib="CITB-HS-A"

misc_feature

1. 138

/note="match to EST AA643257 (NID:g2568475) nr61h09.s1"

misc_feature

1. 338

/note="match to EST D44965 (NID:g1572440)"

```
misc_feature 42. .118
/notes="similar to EST D44969 (NID:g1572444)"
STS complement(106. .396)
/db.xref="GI:3800554"
misc_feature 119. .483
/notes="similar to EST AW236834 (NID:g6569223) xm49f08.xl"
misc_feature 119. .483
/notes="similar to EST AI609282 (NID:g4618449) tw83h06.xl"
misc_feature 119. .378
/notes="similar to EST D44969 (NID:g1572444)"
misc_feature 146. .483
/notes="similar to EST N80671 (NID:g1243372) zb03a07.sl"
misc_feature 164. .483
/notes="similar to EST AW204636 (NID:g6504108)"
misc_feature 204. .206
/notes="match to EST AA643257 (NID:g2568475) nr61h09.sl"
misc_feature 306. .652
/notes="match to EST AA804480 (NID:g2875993) n233c09.sl"
misc_feature 335. .520
/notes="similar to EST AW177547 (NID:g6443584)"
misc_feature 577. .666
/notes="similar to EST AW177547 (NID:g6443584)"
repeat_region 595. .711
/rpt_family="Alu"
repeat_region 712. .740
/rpt_family="(CAAAA)n"
repeat_region 741. .898
/rpt_family="Alu"
repeat_region 899. .919
/rpt_family="AT-rich"
repeat_region 965. .1106
/rpt_family="Alu"
repeat_region 1107. .1395
/rpt_family="Alu"
repeat_region 1396. .1570
/rpt_family="Alu"
repeat_region 1571. .1662
/rpt_family="L1"
repeat_region 1665. .1817
/rpt_family="L1"
repeat_region 1834. .1856
/rpt_family="L2"
repeat_region 1888. .2189
/rpt_family="Alu"
repeat_region 2256. .2546
/rpt_family="Alu"
repeat_region 3070. .3391
/rpt_family="Alu"
repeat_region 3493. .3780
/rpt_family="Alu"
repeat_region 3792. .4101
/rpt_family="Alu"
repeat_region 4108. .4232
/rpt_family="Alu"
repeat_region 4247. .4422
/rpt_family="Alu"
repeat_region 4423. .4729
/rpt_family="Alu"
repeat_region 4730. .4866
/rpt_family="Alu"
repeat_region 4941. .5099
/rpt_family="L1"
repeat_region 5222. .5527
/rpt_family="Alu"
repeat_region 5812. .5883
/rpt_family="MERL_type"
repeat_region 5884. .6194
/rpt_family="Alu"
repeat_region 6207. .6397
/rpt_family="MaLR"
repeat_region 6398. .6469
/rpt_family="MaLR"
repeat_region 6470. .6771
```

```
repeat_region /rpt_family="Alu"
6772. .7113
/rpt_family="MaLR"
7114. .7422
/rpt_family="Alu"
7423. .7468
/rpt_family="MaLR"
7469. .7644
/rpt_family="MaLR"
7645. .7774
/rpt_family="MERL_type"
7860. .8155
/rpt_family="Alu"
8194. .8466
/rpt_family="Alu"
8479. .8599
/rpt_family="MIR"
8599. .8600
/notes="similar to EST W78139 (NID:g273876)"
repeat_region 8601. .8916
/rpt_family="Alu"
8787. .8803
/notes="similar to EST AA463503 (NID:g2188387) zx72c10.sl"
misc_feature 8810. .9112

Query Match 65.4%; Score 1735.2; DB 9; Length 124526;
Best Local Similarity 74.7%; Pred. No. 0;
Matches 1831; Conservative 0; Mismatches 565; Indels 54; Gaps 4;

Qy 1 t g t c t a g g c c a t a g c t t g c g a g t g c c g g g g c g g g g c t c t c a g c c t g c a g a g a g g c a g 60
Db 122131 T G T C T A G G C C A T A G C T T G G C A G T G C C G G G G G G G C T C T C A G C C T G G C A G A G G C A G 122190

Qy 61 g a c c t c a g g g a a a g g g c t a g a c g c c t g c c g c g t g t g g g c t g g g c t g c a c a g g g g 120
Db 122191 G A C C C T C A G G G G A A A G G G G C T G G A C G C G C T G C C G G G G T G T G G G C T G C A C G G G G 122250

Qy 121 c g g a a g a a a g c g c a t g c c c g g g g c t t t g g g a t g g g c a g t c c a g g g g g g c t c c c c g 180
Db 122251 C G G A A G A A A G C G G C A T G C C C G G G G C T T T G G G A T G G C A G T C C A G G G G G C T C C C G 122310

Qy 181 g a g a g g g g a c a g a c c a a g g c t g t g a g g g g c g t g g a a a a c c g c c c a g g c t c t g c t 240
Db 122311 G A G A G G G G A C A C A G A C C A A G G C T G T G A G G G G C G T G G A A A A C C G C C A C A G G C T C G C T 122370

Qy 241 g c a g g c a a g g t c c t g t a c g g g g c a g c g c c t t g t c c c c g c g g g t c g t g c 300
Db 122371 G C A G G C A A G G G T C C T T G T G T G A C G G G G C A G C C G C C T T T G T C C G C G G G G T C G T G C 122430

Qy 301 a g a c t a c c g g c c c c t a c t g c c c c c a c t t c c t c g a c c a g g g t g c c a t c t g a g t c c c 360
Db 122431 A G A C T A C C G G C C C C T A C T G C C C C C A C T T C C T C G A C C A G G G G T G C C A C T G A G T C C C 122490

Qy 361 t g g g g c a g g g g c c c t c g g g c t t t a c a g a c g c c c c g t c c c g c t g g g c a g g t c g t c c a 420
Db 122491 T G G G G C A G G G G C C C T C G G G C T T T T A C A G C G C C C G T C C C G C T G G G C A G G T C G T C C A 122550

Qy 421 t c c g a c g c a c a g a c a t c a c a g c g g t c g c g a a g c c c t c a g c c a g a g c g c t a t c 480
Db 122551 T C C G C A A C G T G C A C A G C A T C C A C C A G C G G T C G C G A A G G C C T C A G C C A G G A C G C T A T C 122610

Qy 481 g c c g c a a c a g c g t c c g t t t t c t g c a g c a g c g c g c c a g c g c g c g c g g c g g a c c g c a g a 540
Db 122611 G C C G C A A C A G C G T C C G T T T T T T G C A G A C G C A G C G C C G C C A G G C G G C G G A C C G C A G A 122670

Qy 541 g c c c g g a g c c c g c t c g g t a g t c a g c g g g a g a g g g c a g a a g g a a g c c c t a g g g g 600
Db 122671 G C C C G G A G C C C G C T C G T G A G T C A G C G G A G A G G G C A G A A G G C A A G C C C T A G G G G 122730

Qy 601 c g g a g t c a g c g g a g a g c g g g c c a g a g g c c a g a g t a g c g g g c g g a g a g a g a c c a g a g 660
Db 122731 C G G A G T C A G C G G A G A G C G G C C A G A G G T A G G C C C A G A G T A G G C C C A G A G T A G G C G G G A C C A G A G 122790
```

QY 661 ggccgaatcagagagagcgaggactggagcggggagcgagagagcgagcgctag 720
|||||
Db 122791 GCGGAATCAGAGGAGAGCGGGACTTGGAGCGGGGCACAGGAGGAGCCAGCGCTAG 122850
QY 721 gggcgagcgatccctaaagagcgagtcagagagagagccacaagcgaggcgagc 780
|||||
Db 122851 GGGCGGAGCGATCCCTAAGAGCGGAGTCAAGAGGAGAGGCACAGCGGAGGAGGCG 122910
QY 781 cagagcgagagcagagtgtagaccgcgcgggcgagggcgagagcagagcgctgagcg 840
|||||
Db 122911 CAGACGGGACAGAGTTGGAGACCGCGCGGGCGGAGGCCAGAGCGCTGTGGCG 122970
QY 841 gggccagtgcgggcggggctgactcgcccgcccgctctctcccgcaagagagag 900
|||||
Db 122971 GGGCCAGTGTGCGGGCGGGGCTGTACTCGGCCCGCTCTCTGCCGAGAGGAG 123030
QY 901 cggagagcgagcgctctaaaccgagcgcgcggtgcccccgcgagcgccgacetc 960
|||||
Db 123031 CGGCAGACGACGCGCTAAACCGAGCGGCGGTGCCCGCGCGGCGGCGGCGGACCTC 123090
QY 961 atcctgaaccgctgcagcgagagcaccagcggaagctggcgctcgccctgagagctag 1020
|||||
Db 123091 ATCTGTAAACCGTGCAGGAGAGACCAAGCGAAGCTGGCGTCTGCGCTGTATACGTTCT 123150
QY 1021 agcgagtcgccagctagctcgccgcttgcgcgcccgcgtgcgtgtatatacgtgtct 1080
|||||
Db 123151 AGCGCAGTCCCGAGCTAGCGTCTCGGCCCTTGCCGCCCGCTGCTGTATACGTTCT 123210
QY 1081 atagagcctggctgcagcgagagcgagcccgagcccgagccctgtccagcgcgctccgc 1140
|||||
Db 123211 ATAGAGCTGGCGTGCAGCGCGAGGCGAGCCCGAGCCCTGTGCAGCGGCTCCCGC 123270
QY 1141 caccctcaataatgttcttgagtgagcgagcgctgcaggaatcgagggagggcg 1200
|||||
Db 123271 CACCCTCAATAATGTTGCTTGGAGTGCACCGAGCTCTGCAGGAATCGAGGAGGCGC 123330
QY 1201 ggtccgcccaggggtatttctaaagtgagagcagagaggtgtgagttctgtnnnnn 1260
|||||
Db 123331 GGCTCGGCCAGGGTTA-TTTCTAAGTTGAGGACA-GGAGGTTGTGAGTTCTG 123382
QY 1261 nnn 1320
Db 123383 -----CTGGGGGAAAGTTGC 123397
QY 1321 nnn 1380
Db 123398 AAGAGCCGAGGTCTGGTTGCATGTTGCCCTTGCTTGCCCAAGACAGGTTTGCACAAG 123457
QY 1381 nnn 1440
Db 123458 CCAAGTTCAAGAGGAACCTCCCGGTTCTCTGCTGACCGTTTGGTGCAGAAACCACTGCTTG 123517
QY 1441 nnn 1500
Db 123518 GACTCTGCGGNAGAGTCTCTGAGATGGGTGCACACACTGCAGAGGCGACCCCTGTCTC 123577
QY 1501 nnn 1560
Db 123578 ATGGCAGGAGACAGGCTGCCGCTCCAGGGTGTAGGAGTGACCTCATAGCTGGGATAAAAA 123637
QY 1561 nnn 1620
Db 123638 TATATATAACTTAGTTTCGGCGCGGTGGGTCAAGCTGTAAACCCAGCGCTTTGGGAGA 123697
QY 1621 nnn 1680
Db 123698 CCGAGGTGGAGATCCGTTGAGCTCAGAGTTCGAGACCGCTGGCCACATGCTGAA 123757
QY 1681 nnn 1740
Db 123758 ACCCCATCTACCAAAAAATATAAAATATAGCTGGCGGTGGTGGCATGCCATCCATAATCC 123817
QY 1741 nnn 1800
;

Db 123818 CAGCTACTGGGAGGCTGAGGCATGAGAAATCGCTTGAACCGGGAGGCAGATGTTGCACT 123877
QY 1801 nnn 1860
Db 123878 GAGCCGAGACGCGCCACTTGGACTCCAGCTGGACAACATGTGAACCCCATCTGTACC 123937
QY 1861 nnnnnnnntaaatagctggcgctggcgatgcacacatccagactactggagag 1920
|||||
Db 123938 AAAATATAAAATTTAGCTGGCGGTGGTGGCATGCATCCACAATCCAGCTACTTGGGAG 123997
QY 1921 gctgaggcatgagaatcgctgaaccgggagcgagatgttcagtgagccgagagcgcg 1980
|||||
Db 123998 GCTGAGGCATGAGATCGTTGAACCGGGAGCGAGATGTTGCAGTGAGCCGAGAGGCG 124057
QY 1981 ccactgcaactcagcctggactacagagcgagactctatctcaaaaaaataaaaaa 2040
|||||
Db 124058 CCACCTGCATCCAGCTGGACTACAGAGCGAGACTCCATCTC-AAAAAAATAAAAAA 124116
QY 2041 aaagtaacttagtgaggggtgctctgttattcactgagaccgtgcccggttatgag 2100
|||||
Db 124117 AAAGTAACCTTAGTGCAGGCTGCTCTCTGTTATTCATGAGACCGTGCCTGTTATGAG 124176
QY 2101 gttgtaccagaagaagtattcactatgcacactattcaccgctcacccctagcattgaa 2160
|||||
Db 124177 GTTGTACCAGAAAGCAAGTATTCACTATGCACACTATTCAACCGCTCACCTAGCATTGAA 124236
QY 2161 gccagcctgtagctgaaagcccttgccttggagcgagcttctcccaaaatgcagaca 2220
|||||
Db 124237 GCCAGCCTGTAGCCTGAAAGCCTTTGCTTTGAGGCGAGGCTTTCCCAAAATGAGACA 124296
QY 2221 cgaagtgcaaatgaagctgcagctcctctgcaaaagatgtaactgtcacgaagccagc 2280
|||||
Db 124297 CGAAGGTGCAAGTCAAGCTGCCAGTCTTGTGCAAAAGATGTAACTTGTACGAAGGCCAGC 124356
QY 2281 agtgcagggagagctgtccacatttcgggaagtggctatgtgagagcgggggagcg 2340
|||||
Db 124357 AGTGCAGGGAGAGCTGCCACATTTGCGAAGTGGCTATGTGAGGACGGGGAGGCGG 124416
QY 2341 gtccttagaataagacaatacataaaggagagatatcagaaaaatcgtaagggagc 2400
|||||
Db 124417 GTCCCTTAGAGATAAGACAATCATAGGGGAGATATCAGAGAAATCGTAAGGGGAGC 124476
QY 2401 agatggttgcaagaagaatagctgaccatcgcaaggaactgccaagctt 2450
|||||
Db 124477 AGATGTTGTCAAGAGATAGGCTGACCATCGAAGGACTGGCAGAAGCTT 124526

RESULT 6

AF184614 AF184614 17302 bp DNA PRI 07-MAY-2001
LOCUS Homo sapiens p47-phox (NCF1) gene, complete cds.
DEFINITION AF184614
ACCESSION AF184614
VERSION AF184614.1 GI:6983939
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS 1 (bases 1 to 17302)
Chanock, S.J., Roesler, J., Zhan, S., Hopkins, P., Lee, P.,
Barrett, D.T., Christensen, B.L., Curnutte, J.T. and Gorlach, A.
TITLE Genomic structure of the human p47-phox (NCF1) gene
JOURNAL Blood Cells Mol. Dis. 26 (1), 37-46 (2000)
MEDLINE 20238075
PUBMED 10772875
REFERENCE 2 (bases 1 to 17302)
Chanock, S.J., Roesler, J., Zhan, S., Hopkins, P., Lee, P., Barrett, D.,
Christensen, B.L., Curnutte, J.T. and Gorlach, A.
AUTHORS Direct Submission
TITLE Submitted (08-SEP-1999) Pediatric Oncology Branch, NCI,
JOURNAL NIH/10/13N240, Bethesda, MD 20892, USA
FEATURES Location/Qualifiers

Scripps Research Institute, 10550 N. Torrey Pines Rd., CAL 1, La Jolla, CA 92037, USA

Location/Qualifiers

1..1645

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="p1 clone p43"

/chromosome="7"

/map="7q11.23"

/note="Human genomic p1 clone p43 is a p47-phox pseudogene which is defined by the presence of a GT deletion at the beginning of exon 2"

Join(U60970.1:<19..7258,U60971.1:1..3348,1..1619)

/gene="NCF1"

/pseudo

398..502

/gene="NCF1"

/pseudo

/number=9

974..1119

/gene="NCF1"

/pseudo

/number=10

1454..1619

/gene="NCF1"

/pseudo

/number=11

293 a 504 c 610 g 239 t

BASE COUNT

ORIGIN

Query Match 39.7%; Score 1052.4; DB 9; Length 1646;

Best Local Similarity 99.3%; Pred. No. 1.5e-196;

Matches 1078; Conservative 0; Mismatches 6; Indels 2; Gaps 2;

Qy 1 tgtctagggccatagcttggcagtgccggggggggggtctcagccttgccagggagggcag 60
Db 562 TGTCTAGGCGCATAGTTGGCAGTGGCGGGGGGGGGCTCTCAGCCTGGCAGGAGGCGAG 621
Qy 61 gaccttcaggggaaagggctgaacgcctgcccgcgtgtaggggtgtaggggagggg 120
Db 622 GACCCCTACGGGGGAAAGGGGCTGACGCGCCTGGCGGGGGTGGGGCTGGCAGGGGG 681
Qy 121 cggaaaggaaagcggcagtgccgggggggtttggggatggcagtcaggggggtccccc 180
Db 682 CGGAAGAAAGCGCGATGCCCGGGGGCTTTGGGATGGGCAGTCCAGGGGGCTCCCGG 741
Qy 181 gaggggggacacagaccgaaggtggtgagggcgtagaaacccgccagggctctgct 240
Db 742 GAGAGGGGACACAGACCGAAGGCTGTGTAGGGGGCGTGGAAACCGCCAGGCTCTGCT 801
Qy 241 gcagggcaagggtccttgcgtgacgggggagccgcctcttgcctcccggggtcgtagc 300
Db 802 GCAGGGCAAGGGTCTCTGTGTGACGGGGGACGCGCTTGTCCCGGGGGTGGTGC 861
Qy 301 agactaccggccccctactgcccccaactctctcagaccaggggtgcccattctgagtc 360
Db 862 AGACTACCGGGCCCCCTACTGCCCCCCACTTCTCGACACAGGGGTGCCATCTCAGTCC 921
Qy 361 tgggggagggggccctcctgggctttgacgacgccccctcccgctggggccaggtcgcca 420
Db 922 TGGGGGCAAGGGGCGCCCTGCGGGCTTTGACGACGCGCCCTCCCGCTGGGCGAGTCTCA 981
Qy 421 tccgaacgcgcacagcatccacagcaggtgcggaagcgtcagcagcagcgtatc 480
Db 982 TCCGCAACGTGCACAGCATCCACAGCGGTGCGGAAGCGCTCAGCCAGGACGCTATC 1041
Qy 481 gccgaacagcgtccgtttcttcagcagcagcagcgcgcagcgcgcgcgcgcgcgcaga 540
Db 1042 GCCGCAACAGCGTCGCTTTCTGACAGCAGCGCCGCGGCGGGGCGGACCGCAGCA 1101
Qy 541 gccccggagccgcctcgtgtagtcagcgggaggggcagaaaggcagcccttagggg 600

Db 1102 GCCCCGGGAGCCCGCTGGTGAAGTGCAGC-GGAGAGGGCAGGGAAGGCCCTAGGGG 1160
Qy 601 cggagtcacggagagagcggggccagagcagggccagagtagcggggcgaggaccagag 660
Db 1161 CGGAGTCACGGGAGAGCGGGGCCAGAGGTAGGGCCAGAGTAGCGGGCGGAGCCAGAG 1220
Qy 661 ggcgggaatcagagggagagcgggggactgagagcgggggcagagggagggagcagcg-c 719
Db 1221 GCGGGAATCAGAGGAGAGCGCGGGGACTGGAGCGGGGGCAGAGGAGGAGCCAGCC 1280
Qy 720 gggggcgagcgcctcctaagagggcgagtcagagggagagggcacaagcggggagcgag 779
Db 1281 GGGGGCGGAGCATCCTTAAGAGCGGAGTCAAGAGGAGAGGACACAAGCGGAGCGGAG 1340
Qy 780 ccagagcggcagcagcagagagttgagacccgcggcgggggcagggccagagcgtgtgggc 839
Db 1341 CCAGAGCGCGGAGCAGGAGTTGGAGACCGCGGGCGGGCCAGGAGCAGAGCGCTGTG 1400
Qy 840 gggggcagtgctgcggggcgggcgctgctgactcgggccccctctctctcccgccagagga 899
Db 1401 GGGGCCAGTGTGGGGGGCGGGCGTCTGACTCGGCCCGCTCTCTGCCCGCAGAGGAG 1460
Qy 900 gcggcagcgcgcgtctctaaaccgcagcggcggtgtcccccgcggccgagcgccagcct 959
Db 1461 GCGGCAGACGACGCGCTCTAAACCGCAGCGCGGCTGCCCGCGCGGCGGCGGCGG 1520
Qy 960 catcctgaacctgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1019
Db 1521 CATCTTGAACCGCTGCAGCGAGAGCACCAGCGGAAAGCTGGCGGTCTGCGGCTGTG 1580
Qy 1020 gagcgagtcctccagcagtcagcgtctcgcccttgcggccccgcgtgtatatacgtgtc 1079
Db 1581 GAGCGAGTCCCGCAGCTAGCGTCTCGGCCCTTGGCCCGCCGTCCTGTACATACGTG 1640
Qy 1080 tataga 1085
Db 1641 TATAGA 1646

RESULT 8
AC092405/c

LOCUS

DEFINITION

AC092405

AC092405

AC092405.1

HTG; HTGS_PHASE1; HTGS_DRAFT.

KEYWORDS

olive baboon.

ORGANISM

Papio cynocephalus anubis

Eukaryota; Metazoa; Chordata; Mammalia; Eutheria; Primates; Catarrhini; Cercopitheciinae; Papio.

REFERENCE

1 (bases 1 to 184558)

Green,E.D.

Direct Submission

Submitted (04-JUL-2001) NIH Intramural Sequencing Center, 8717

Genome Center

Center: NIH Intramural Sequencing Center

Center code: NISC

Web site: <http://www.nisc.nih.gov>

Contact: nisc_mouse@nhgri.nih.gov

Unpublished

2 (bases 1 to 184558)

Green,E.D.

Direct Submission

Submitted (04-JUL-2001) NIH Intramural Sequencing Center, 8717

Genome Center

Center: NIH Intramural Sequencing Center

Center code: NISC

Web site: <http://www.nisc.nih.gov>

Contact: nisc_mouse@nhgri.nih.gov

Unpublished

2 (bases 1 to 184558)

Green,E.D.

Direct Submission

Submitted (04-JUL-2001) NIH Intramural Sequencing Center, 8717

Genome Center

Center: NIH Intramural Sequencing Center

Center code: NISC

Web site: <http://www.nisc.nih.gov>

Contact: nisc_mouse@nhgri.nih.gov


```

----- Project Information
Center project name: ccx
Center clone name: 170F23
----- Summary Statistics

Sequencing vector: plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 178116 bases at least Q40
Consensus quality: 178998 bases at least Q30
Consensus quality: 181141 bases at least Q20
Insert size: 168000; agarose-fp
Insert size: 184058; sum-of-contigs
Quality coverage: 10.5x in Q20 bases; agarose-fp
Quality coverage: 9.60x in Q20 bases; sum-of-contigs

```

* NOTE: This is a 'working draft' sequence. It currently
* consists of 6 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 13828: contig of 13828 bp in length
 * 13829 13928: gap of unknown length
 * 13929 13157: contig of 17329 bp in length
 * 13158 13157: gap of unknown length
 * 13158 42795: contig of 11438 bp in length
 * 13158 42795: gap of unknown length
 * 42796 61486: contig of 18591 bp in length
 * 42796 61487: gap of unknown length
 * 61487 111399: contig of 50053 bp in length
 * 61587 111640 111739: gap of unknown length
 * 111640 184558: contig of 72819 bp in length.
 * 111740

FEATURES	Location/Qualifiers
1111740	164338: CONCISE

```

location/Quadrifidius
1. .184558
/organism="Papio cynocephalus anubis"
/db_xref="taxon:955"
/clone="RP41-170F23"
/clone lib="RP41"

```

misc_feature

ATTORNEY GENERAL

misc_feature

1

misc_feature

misc_feature

2017-05-11

misc_feature

1

misc_feature

COUNT 4818

IN
COUNTY

11

ery Match

st Local Simil

atches 1570; 0

161 catccan

101 cagucca
111

38562 CAAGCGC

2
2
2
2
1
5
2
2
2
2

221 aaaccg

—

38502 GGGGCGC

18C +++++

287 ttgtccc

[illegible]

```

----- Project Information
Center project name: cck
Center clone name: 170F23
----- Summary Statistics
Sequencing vector: plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 178116 bases at least Q40
Consensus quality: 179898 bases at least Q30
Consensus quality: 181141 bases at least Q20
Insert size: 168000; agarose-1p
Insert size: 184058; sum-of-contigs
Quality coverage: 10.52x in Q20 bases; agarose-1p
Quality coverage: 9.60x in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 6 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 13828: contig of 13828 bp in length
* 13829 13928: gap of unknown length
* 13929 31257: contig of 17329 bp in length
* 31258 31357: gap of unknown length
* 31358 42795: contig of 11438 bp in length
* 42796 42895: gap of unknown length
* 42896 61486: contig of 18591 bp in length
* 61487 61587: gap of unknown length
* 61588 111639: contig of 50053 bp in length
* 111640 111739: gap of unknown length
* 111740 184558: contig of 72819 bp in length.
* Location/Qualifiers
* 1. 184558
* /organism="Papio cynocephalus anubis"
* /db_xref="taxon:9555"
* /clone="Rp41-170F23"
* /clone_lib="Rp41"
*
* 1 13828
* /note="assembly_fragment"
* 13929. .31257
* /note="assembly_fragment"
* 31358. .42795
* /note="assembly_fragment"
* clone_end:T7
* vector_side:left"
* 42896. .61486
* /note="assembly_fragment"
* 61587. .111639
* /note="assembly_fragment"
* 111740. 184558
* /note="assembly_fragment"
* clone_end:SP6
* vector_side:right"
*
* BASE COUNT 48162 a 44484 c 44779 g 46622 t 511 others
* ORIGIN
*
* Query Match 39.2%; Score 1039.6; DB 2; Length 184558;
* Best Local Similarity 62.9%; Pred. No. 1.3e-194;
* Matches 1570; Conservative 0; Mismatches 629; Indels 296; Gaps 10;
*
* QY 161 cagtcacaggggctccccagagaggggacagaccagagctggtgagggcgctgg 220
* || | |||| | || | || | || | || | || | || | || | || | || | ||
* Db 138562 CAAGCGCGGGCGCGCGCCCGCAGGTAGCGGGGTTCCTCGGGGCTGGCGAGG 138503
*
* QY 221 aaacccgcccagctgctgtcagggcaagggctcctgtgtcgtgacggggcgagccgctc 280
* || | |||| | || | || | || | || | || | || | || | || | || | ||
* Db 138502 GGGCGGCACACAGGTTTCGCTCTCTAGGCAGCTTGGCAGTGCCTGGGGCGGGGGCTCTC 138443
*
* QY 281 ttgtcccgccgggtgctgtgcagactacgcggcccccactgctgcccccacttctcggacca 340

```


Query Match		38.8%;	Score 1028.6;	DB 9;	Length 1619;
Best Local Similarity		99.4%;	Pred. No. 6.8e-192;		
Matches 1053;		Conservative	0;	Mismatches	4;
				Indels	2;
				Gaps	2;
Qy	1	tgtctagccatagcttgccagtcgagcgccgaggggctctcagctgagagagcag	60		
Db	562	TGCTAGGCATAGCTTGGCAGTCCCGGGCGGGGGCTCTCAGCCTGGCAGGAGGCGAG	621		
Qy	61	gacctcacggggaaggggctgagcgccttgccgcggtgtggggctggcagcgggg	120		
Db	622	GACCTCACGGGGAAGGGCTGGACGCGCTTGCCCGCGGTGTGGGCTGCACAGGGG	681		
Qy	121	cgaagaaagcgcgatgccccggggctttgggagtgggcagtcacgggggctccccg	180		
Db	682	CGAAGAAAGCGCGATGCCCCGGGGCTTTGGGGATGGGAGTCCAGAGGGGGTCCCG	741		
Qy	181	gagaggggagcagacagacgaagctgtagggcgctggaagcctggaaccccgagctgct	240		
Db	742	GAGAGGGGAGCAGACGACCAAGGCTGGTGAGGGGGCTGGAACCCGCCAGGCTCTGCT	801		
Qy	241	gcagggaaggtctctgtgtagcggggagcgcctcttctccgcgaggggtctgctgc	300		
Db	802	GCAGGGCAAGGCTCTTGTCTGACGGGGGCGCGCGCTTCTGTCCCGCGGGTCTGTC	861		
Qy	301	agactaccggccccctactgccccccacttctctcgaccaggggtgccccatctgagtc	360		
Db	862	AGACTACGGCCCCCTACTGCCCCCACTTCTCTCGACGAGGGGTGCCATCTGAGTCCC	921		
Qy	361	tggggcagggcgccctgggttttaagacgcgcctcccgctggccaggtcgtcca	420		
Db	922	TGGGGGAGGGGGCCCTCGGGCTTTGACGACGCCCTCCCGCTGGCGCAGGTCTCCA	981		
Qy	421	tccgaacgcgcacagcatccaccagcggtctcggaagcgctcagccagagacccatc	480		
Db	982	TCCGCAAGTGCACAGCATCACACAGCGGTGCGGAGAGCGGCTCAGCCAGGACGCTATC	1041		
Qy	481	gcgcgaacagctcgtttttctgacgagcagcagcccgccagggcgccgggacccgaga	540		
Db	1042	CCCCAAGCGGCTCGGTTTCTGCAGCAGGCGACGCCCGCAGCGCGCGGACCGCAGA	1101		
Qy	541	gccccgaagcccgctcggtgaagtcagcgaggaagcgaggaagcaagccctaggg	600		
Db	1102	GCCCCGGAGCCGCTCGGTGAGTGCAGC-GGAGAGGCGAGGAAGGCAAGCCGTAGGGG	1160		
Qy	601	cggagtcagcgagagcggggcccagagcgagggccagagtagcgggcggggaccagag	660		
Db	1161	CGGAGTCAGCGGAGAGCGGGGCCAGAGGTAGGCCAGAGTACGGGGCGGGACAGAG	1220		
Qy	661	ggcggaatcagagggagcggggactgagcgggggcagagagagagagcagcgc-ta	719		
Db	1221	GGCGGAATCAGAGGGAGAGCGGGGACTTGGAGCGGGGCGGAGAGGAGAGCCAGCGCTTA	1280		
Qy	720	ggggcgagcgatccctaagagcgagtcagagggagaggaagcaagcggggagcgagg	779		
Db	1281	GGGGCGGAGCGGATCCCTAAGAGCGGAGTTCAGAGGAGAGGACAAAGCGGGAGCGGAGG	1340		
Qy	780	ccagagcgcgagcagagtggtgagacgcgcggcgggcgagggcagagagcgctgtgggc	839		
Db	1341	CCAGAGCGCGAGGAGGAGTGGAGACCGCGCGGGCGGAGGCGGAGAGGCGCTGTGGGC	1400		
Qy	840	ggggccaggtgtgcggggcggggctctgactcggccccctctctgccccagagagga	899		
Db	1401	GGGGCAGTGTGCGGGCGGGGCTCTGACTCGGCGCGGCTCTCTGTCGCCCGCAGAGGAGA	1460		
Qy	900	gcggcagacgagcgctctaaacgcagcgcggtgtcccccccgcgcgagcgcgacct	959		
Db	1461	CGGCAGACGAGCGGCTCTAAACCGCAGCGCGGCTGCCCCCGCGCGGAGCGCGACCT	1520		
Qy	960	catctgaaccgctgcagcgagagacaccaagcgaagctgctgtcgtctgagcgctg	1019		
Db	1521	CATCTGAACCGCTGCAGCGAGACGACCAAGCGGAAGCTGCGCTGTGCCGTCTGAGGCTG	1580		
Qy	1020	gagcgagtcgccagtagcgtctcggcccttgcgcgcc	1058		

Db	1581	GAGCGAGTCCCCAGCTAGCGTCTCGGCCCTTGGCGCCC				1619
RESULT	10					
HS47P41S04	1619 bp	DNA	PRI	07-JAN-1998		
LOCUS	Homo sapiens p47-phox pseudogene, clone p41, exons 9-11.					
DEFINITION						
ACCESSION	U61244					
VERSION	U61244.1	GI:2754728				
KEYWORDS						
SEGMENT	4 of 4					
SOURCE	human.					
ORGANISM	Homo sapiens					
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.						
REFERENCE	1 (bases 1 to 1619)					
AUTHORS	Gorlach,A., Lee,P.L., Roesler,J., Hopkins,P.J., Christensen,B., Green,E.D., Chanock,S.J. and Curnutte,J.T.					
TITLE	A p47-phox pseudogene carries the most common mutation causing p47-phox- deficient chronic granulomatous disease					
JOURNAL	J. Clin. Invest. 100 (8), 1907-1918 (1997)					
MEDLINE	97474758					
REFERENCE	2 (bases 1 to 1619)					
AUTHORS	Chanock,S.J., Roesler,J., Hopkins,P.J., Lee,P.L., Bassett,D.T., Christensen,B., Curnutte,J.T. and Gorlach,A.					
TITLE	Characterization of the genomic structure of the p47-phox gene					
JOURNAL	Unpublished					
REFERENCE	3 (bases 1 to 1619)					
AUTHORS	Gorlach,A., Lee,P.L., Roesler,J., Christensen,B., Chanock,S.J. and Curnutte,J.T.					
TITLE	Direct Submission					
JOURNAL	Submitted (19-JUN-1996) Molecular and Experimental Medicine, The Scripps Research Institute, 10550 N. Torrey Pines Rd., CAL 1, La Jolla, CA 92037, USA					
FEATURES	Location/Qualifiers					
source	1..1619					
	/organism="Homo sapiens"					
	/db_xref="taxon:9606"					
	/clone="p41"					
	/chromosome="7"					
	/map="7q11.23"					
	/note="Human genomic P1 clone p41 is a p47-phox pseudogene which is defined by the presence of a GT deletion at the beginning of exon 2."					
exon	398..502					
	/pseudo					
	/number=9					
exon	974..1119					
	/pseudo					
	/number=10					
exon	1454..1619					
	/pseudo					
	/number=11					
BASE COUNT	287 a 498 c 604 g 230 t					
ORIGIN						
Query Match						
Best Local Similarity		38.7%;	Score 1027;	DB 9;	Length 1619;	
Matches 1052;		Conservative	0;	Mismatches	5;	
				Indels	2;	
				Gaps	2;	
Qy	1	tgtctagccatagcttgccagtcgagcgccgagggctctcagcctggcagagagcag	60			
Db	562	TGCTAGGCCATAGTTTGGCAGTCCGGGGCGGGGGCTCTCAGCCTGGCAGGAGGCGAG	621			
Qy	61	gacctcacggggaaaggggctgagcgcgctggccgcggtgtggggctggcagcgaggg	120			
Db	622	GACCTCACGGGGAAAGGGGCTGGACGCGCTGCGCGCGGTGTGGGCTGCGACGGGG	681			
Qy	121	cgaagaaagcgcgatgcccgggggttttgggagtgggcagtcacggggggtcccg	180			
Db	682	CGAAGAAAGCGCGATGCCCGGGGCTTTGGGGATGGGAGTCCAGGGGGGTCCCG	741			

QY 181 gagaggggagacagacagcgaagcgtggtgagggcgctggaaacccgccagcgtctgct 240
|||||
Db 742 GAGAGGGGACACAGACCGAAGGCTGCTGAGGGCGTGGAAACCGCCAGGCTCTGCT 801

QY 241 gcagggaagggctcttctgctgacggggcagccgcctcttctgcccgcgggtcgtgc 300
|||||
Db 802 GCAGGGCAAGGCTCTGCTGCTGACGGGGCAGCGGCTCTGTGCCGCGGGGTCGTGC 861

QY 301 agactaccggccctactgcccacactctctcgacacaggggtgcccattcgaagtc 360
|||||
Db 862 AGACTACCGGCCCCCTACTGCCCCCACTTCTCTCGACAGGGGTGCCATCTGAGTCCC 921

QY 361 tggggcgaggcgccctcggttctgacgcgccccctcccgctggcgaggtcgtcca 420
|||||
Db 922 TGGGGGACGGGCGCCCTCGGGCTTTGACGAGCGCCCCCTCCGCTGGCCAGGCTGCTCA 981

QY 421 tcgcgaacgcgcacagcatccaccagcgtcgcggaagcgcctcagccagcagcctatc 480
|||||
Db 982 TCCGCAAGCTGCACAGCATCCACCAAGCGTCCGGAAGCGGCTCAGCCAGGACGCTATC 1041

QY 481 gccgaacagcgtccgtttctgacagcagcagccgcgcagggcgccgggaccgcaga 540
|||||
Db 1042 GCCGCAACAGGCTCCGCTTTCTGACGACGACGCGCGCCGCGCGGACCGCAGA 1101

QY 541 gccccggagccgcgtcggtgagtcagcgggagagggcaggaaggcgaagccctagggg 600
|||||
Db 1102 GCCCGGGAGCCCGCTCGGTGAGTGACG- GGAGAGGCGCAGGAGGGCAAGCCGTAGGGG 1160

QY 601 cggagtcagcgggagagcggggccagagggcagggccagagtagcggggcggaccagag 660
|||||
Db 1161 CGGAGTCAGCGGGAGAGCGGGCGGCGAGAGGTAGGCCAGAGTAGCGGGCGGGACACAG 1220

QY 661 ggcgaatcagaggagagcggggactgagggcgggggcagagaggagagcagcgc-ta 719
|||||
Db 1221 GCGGAATCAGAGGAGAGCGGGGACTGGAGCGCGGGGCGCAGAGGAGCAGCGCTTA 1280

QY 720 gggggcgagcgatccctaaagcggagtcagagggagagagcacaagcggagggcagg 779
|||||
Db 1281 GGGGCGGAGCGATCCCTTAAGAGCGGAGTCAGAGGAGAGGCACAAAGCGGAGCGGAGG 1340

QY 780 ccagagcgcgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 839
|||||
Db 1341 CCAGAGCGGAGCAGGAGGAGTGGAGACCGCGCGCGCGCGGCGGAGCGCTGTGGCG 1400

QY 840 ggggcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 899
|||||
Db 1401 GGGGCGAGTGTGGGGGCGGGCGGTCTGACTCGGCGCGCGCTCTCTGCGCGCAGAGGAGGA 1460

QY 900 gcggcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 959
|||||
Db 1461 GCGGCAGACGACGCGCTCTAAACCGCAGCGCGGTGCGCCCGCGCGGAGCGCGGACCT 1520

QY 960 catcctgaacgcgtcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1019
|||||
Db 1521 CATCTGAACCGCTGCAGCGAGAGCACCAGCGGAAGTGGCGTCTGCGGCTGAGGCGTG 1580

QY 1020 gagcgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1058
|||||
Db 1581 GAGCGCAGTCCCAGCTAGCGTCTCGGCGCTTGGCGGCC 1619

RESULT 11
HS47PHOX03
LOCUS HS47PHOX03 1645 bp DNA PRI 07-JAN-1998
DEFINITION Homo sapiens p47-phox (NCF1) gene, exons 9-11, and complete cds.
ACCESSION U57835
VERSION U57835.1 GI:2754711
KEYWORDS 3 of 3
SEGMENT human.
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1645)
Gorlach,A., Lee,P.L., Roesler,J., Hopkins,J., Hopkins,P.J., Christensen,B.,
Green,B.D., Chanock,S.J. and Curnutte,J.T.
A p47-phox pseudogene carries the most common mutation causing
p47-phox- deficient chronic granulomatous disease
J. Clin. Invest. 100 (8), 1907-1918 (1997)
97474758
2 (bases 1 to 1645)
Chanock,S.J., Roesler,J., Hopkins,J., Lee,P., Bassett,D.T.,
Christensen,B., Curnutte,J.T. and Gorlach,A.
Genomic Structure and the Identification of Multiple Polymorphisms
in the p47-phox Gene
Unpublished
3 (bases 1 to 1645)
Gorlach,A., Roesler,J., Christensen,B., Chanock,S.J. and
Curnutte,J.T.
Direct Submission
Submitted (09-MAY-1996) P. Lee, Molecular and Experimental
Medicine, Scripps Research Institute, 10550 North Torrey Pines Rd.,
La Jolla, CA 92037, USA
Location/Qualifiers
1. 1645
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="7"
/clone="p1 Clone 42"
/map="7q11.23"
Join(U57833.1:<877. 948,U57833.1:4143. 4223,
U57833.1:5957. 6032,U57833.1:6132. 6297,
U57833.1:7657. 7712,U57834.1:507. 629,U57834.1:1092. 1199,
U57834.1:2748. 2865,424. 528,1000. 1145,1480. 1645)
/gene="NCF1"
Join(U57833.1:<877. 8131,U57834.1:1. 3349,1. 1645)
/gene="NCF1"
Join(U57833.1:877. 948,U57833.1:4143. 4223,
U57833.1:5957. 6032,U57833.1:6132. 6297,
U57833.1:7657. 7712,U57834.1:507. 629,U57834.1:1092. 1199,
U57834.1:2748. 2865,424. 528,1000. 1145,1480. 1601)
/gene="NCF1"
/codon_start=1
/product="p47-phox"
/protein_id="AAB95193.1"
/db_xref="GI:2754713"
/translation="MGDTFIRHTALLCFEKFVPSQHYVYVFLVKKWODLSEKVVYRRE
TEYFHKTKEMFPPEAGAINPNRIIPHPAPKWDGQRAAENRGDTLTYEGSTLM
SLPTKISRCPHLLDFFKVRPDDLLKPTDQTKKPTETLYMPKDGKSTADITGPILOS
YRAITANYKRTSGSEMALSTGDVVEVEKSESGWFCQMKARGWIPASFLEPLDSDPE
TEDEPNYAGEPYVAIKAYTAVEGDEVSLLEGEAVEVIHKLLDGMWVIRKDDVTGYPF
SMYLOKSGQDVSOAORIKRGAPRRSSIRNVHSIHORSKRRLSODAYRRNSVRFLOQ
RRQARPGQSPGSPLEERQTORSPQPAVPPPPSADLLINRCSESTKRKLASAV"
424. 528
/gene="NCF1"
/number=9
1000. 1145
/gene="NCF1"
/number=10
1480. 1645
/gene="NCF1"
/number=11
BASE COUNT 291 a 502 c 606 g 246 t
ORIGIN

Query Match 38.6%; Score 1025.4; DB 9; Length 1645;
Best Local Similarity 99.2%; Pred. No. 2.9e-191;
Matches 1051; Conservative 0; Mismatches 6; Indels 2; Gaps 2;

QY 1 tgtctagccatagcttgcagtcgcggcgagggtctctcagcgtgcagagagcagcag 60
|||||
Db 588 TGTCTAGCCCATAGCTTGGCAGTGGCGGGCGGGGGCTCTCAGCCTGGCAGGAGGCAG 647
|||||
QY 61 gaccctcacggggaaaggggctgagcgcgctgagcgcggtgtggtgggctgacgcgggg 120

Db	648	GACCTCACGGGAAAAGGGCTGGACGCGCTGGCCGCGTGTGGGGCTGGCACGGGG	707
Qy	121	cggagaaagcggcgatgccccggggcctttgggagatgggagtcacaggggggctccccg	180
Db	708	CGGAAGAAACGGCGATGCCCGGGGCTTTGGGGATGGGAGTCCACAGGGGGTCCCG	767
Qy	181	gagaggggagacagaccgagctggtgagggcggtggaaaaaccgccaggtctgct	240
Db	768	CAGAGGGGACACAGACCGAAGCTGTGAGGGCGTGGAAAAACCGCCACGGCTGCT	827
Qy	241	gcagggaaggttcctgtgtgacgggggagcgcgcctcttgcgcgcggggtgtgc	300
Db	828	CGAGGGAAGGGTCTTGTCTGTGAGGGGCGAGCGGCTCTTGTCCCGCGGGTCTGTGC	887
Qy	301	agactaccggccccctactgccccccacttctcgtcgaccaggggtgccatctgagtc	360
Db	888	AGACTACCGGCCCCCTACTGCCCCCCCTCTCTCGGACCAGGGTGCCCATCTGAGTCCC	947
Qy	361	tgggggcaggggccccctcgggctttgacgacgcccccgctccgctgggccaaggtcgcca	420
Db	948	TGGGGCAGGGCGCCCTCGCGCTTTGACGACGCCCCCTCCCGTGGGCCAGGTCGTCCA	1007
Qy	421	tcgcgaacgcgcacagatcacacagcggtcgcggaagcgccctcagccagagaccctatc	480
Db	1008	TCCGCAACGTCACAGCATCCACAGCGGTCTCGGAAAGCGCTCAGCCAGGACGCCATC	1067
Qy	481	gccgaacagcgctcgtttctgacgacgacgcgcgcagcgccgagcgacgcgacga	540
Db	1068	CCCCAACAGGCTCGTTCGTTCTGACAGACGACGCGCGCCGCGGACCGCAGCAGAGA	1127
Qy	541	gccccggagccccctcggtgagtgacgagggagagggcaggaagggcaagccctagggg	600
Db	1128	CCCCCGGAGCGCGCTCGTCACTGACGACGCGAGAGGCGAGGCGCAAGCCCTAGAGG	1186
Qy	601	cgagtcagcgagagagcggggcccagagcagcgagggccagagtagcgggcgggaccagag	660
Db	1187	CGGAGTCAGCGGGAGAGCGGGGCGCAGAGGTAGGGCCAGAGTAGCGGGCGGGACCAAG	1246
Qy	661	ggcggaatcagagggagagcggggactggagcgggggcgagagagagcagcgc-ta	719
Db	1247	GCCGATCAGAGGAGAGCGGGGACTGGAGCGGGGTACAGAGGAGGAGCCAGCGCTTA	1306
Qy	720	ggggcgagcgatccctaaagcggagtcagagggagagagcacaagcgagggcaggg	779
Db	1307	GGGGCGAGCGATCCCTAACAGCGGAGTCAGAGGAGAGAGGCACAGCGGAGCGGAGG	1366
Qy	780	ccagagcgcgagcagaggttgagagaccgcgcggcgagggcgagcagagagcgctgtgggc	839
Db	1367	CCAGAGCGGAGGAGGAGTTGGAGACCGCGCGGGCGGCGAGGCCAGAGCGCTGTGGGC	1426
Qy	840	ggggccagtgtcggggcgggcgctgactcggcccccgctctctgcccgcagagagga	899
Db	1427	GGGGCCAGTGTGCGGGCGGGGCGTGTGACTGCGGCCCGCTCTCTGCCCGCAGAGGAGA	1486
Qy	900	gcggcagacgcagcgctctaaacgcagcgcggtgcccccgcgccgagcgccgacct	959
Db	1487	CGGCGACGCGACGCTCTAAACCGCAGCGCGGTGCCCGCGGGCGGACGCCGACCT	1546
Qy	960	cactctgaaccgctgcagcgagagacacaaagcggaagctgctgctgagctgagctg	1019
Db	1547	CATCCTGAACCGCTGCAGCGAGAGCACCAGCGGAAGTGGCGTCTGCGCTCAGGCGTG	1606
Qy	1020	gagcaggtcccaagctagctctgccccttgccgccc	1058
Db	1607	GAGCGAGTCCCCAGCTAGCGTCTCGGCCCCCTTGCGCC	1645

RESULT 12
AC027219 AC027219 149830 bp DNA HTG
LOCUS Homo sapiens clone RP11-729P19, *** SEQUENCING IN PROGRESS ***, 79
DEFINITION unordered pieces.

ACCESSION	AC027219
VERSION	AC027219.3
KEYWORDS	HTG; HTGS_PHASE1.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS	1 (bases 1 to 149830)
TITLE	Birren,B., Linton,L., Nusbaum,C. and Lander,E.
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 149830)
AUTHORS	Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N., Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F., Boguslavsky,L., Bouckhalter,B., Brown,A., Burkett,G., Campoliano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S., Collamore,A., Cooke,P., DeArelano,K., Dewar,K., Diaz,J.S., Dodge,S., Domino,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L., Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J., LaRoque,K., Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Meldrum,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J., Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Olivier,T.M., Oliver,J., Peterson,K., Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo.A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and Zody,M.
TITLE	Direct Submission
JOURNAL	Submitted (28-MAR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT	On Mar 2, 2001 this sequence version replaced gi:11610941. All repeats were identified using RepeatMasker: Smit, A.F.A. & Green, P. (1996-1997) http://ftp.genome.washington.edu/RM/RepeatMasker.html ----- Genome Center Center: Whitehead Institute/ MIT Center for Genome Research Center code: WIBR Web site: http://www-seq.wi.mit.edu Contact: sequence.submissions@genome.wi.mit.edu ----- Project Information Center project name: L8346 Center clone name: 729_P_19 ----- * NOTE: This is a 'working draft' sequence. It currently * consists of 79 contigs. The true order of the pieces * is not known and their order in this sequence record is * arbitrary. Gaps between the contigs are represented as * runs of N, but the exact sizes of the gaps are unknown. * This record will be updated with the finished sequence * as soon as it is available and the accession number will * be preserved. * 1 1569: contig of 1569 bp in length * 1570 1669: gap of 100 bp * 1670 2364: contig of 695 bp in length * 2365 2464: gap of 100 bp * 2465 3099: contig of 635 bp in length * 3100 3199: gap of 100 bp * 3200 3762: contig of 563 bp in length * 3763 3862: gap of 100 bp * 3863 4534: contig of 672 bp in length * 4535 4634: gap of 100 bp * 4635 5474: contig of 840 bp in length * 5475 5574: gap of 100 bp * 5575 6460: contig of 886 bp in length * 6461 6560: gap of 100 bp * 6561 7785: contig of 1225 bp in length * 7786 7885: gap of 100 bp

```

7886      8735: contig of 850 bp in length
8736 8835: gap of 100 bp
8836 9864: contig of 1029 bp in length
9865 9964: gap of 100 bp
9965 10784: contig of 820 bp in length
10785 10884: gap of 100 bp
10885 11653: contig of 769 bp in length
11654 11753: gap of 100 bp
11754 12413: contig of 662 bp in length
12416 12515: gap of 100 bp
12516 13232: contig of 717 bp in length
13233 13332: gap of 100 bp
13333 14178: contig of 846 bp in length
14179 14278: gap of 100 bp
14279 15762: contig of 1484 bp in length
15763 15862: gap of 100 bp
15863 16423: contig of 561 bp in length
16424 16523: gap of 100 bp
16524 17467: contig of 944 bp in length
17468 17567: gap of 100 bp
17568 18533: contig of 966 bp in length
18534 18633: gap of 100 bp
18634 19882: contig of 1249 bp in length
19883 19982: gap of 100 bp
19983 21070: contig of 1088 bp in length
21071 21170: gap of 100 bp
21171 22152: contig of 982 bp in length
22153 22252: gap of 100 bp
22253 23303: contig of 1051 bp in length
23304 23403: gap of 100 bp
23404 24381: contig of 978 bp in length
24382 24481: gap of 100 bp
24482 26183: contig of 1702 bp in length
26184 26283: gap of 100 bp
26284 27663: contig of 1380 bp in length
27664 27763: gap of 100 bp
27764 28566: contig of 803 bp in length
28567 28666: gap of 100 bp
28667 30309: contig of 1643 bp in length
30310 30409: gap of 100 bp
30410 31434: contig of 1025 bp in length
31435 31534: gap of 100 bp
31535 32225: contig of 691 bp in length
32226 32325: gap of 100 bp
32326 33690: contig of 1365 bp in length
33691 33790: gap of 100 bp
33791 34434: contig of 644 bp in length
34435 34534: gap of 100 bp
34535 35776: contig of 1242 bp in length
35777 35876: gap of 100 bp
35877 37225: contig of 1349 bp in length
37226 37325: gap of 100 bp
37326 38564: contig of 1239 bp in length
38565 38664: gap of 100 bp
38665 40341: contig of 1677 bp in length
40342 40441: gap of 100 bp
40442 41665: contig of 1224 bp in length
41666 41765: gap of 100 bp
41766 42998: contig of 1233 bp in length
42999 43098: gap of 100 bp
43099 44123: contig of 1025 bp in length
44124 44223: gap of 100 bp
44224 46278: contig of 2055 bp in length
46279 46378: gap of 100 bp
46379 47723: contig of 1351 bp in length
47730 47829: gap of 100 bp
47830 49626: contig of 1797 bp in length
49627 49726: gap of 100 bp
49727 51471: contig of 1745 bp in length
51472 51571: gap of 100 bp
51572 52901: contig of 1330 bp in length
52902 53001: gap of 100 bp
53002 53821: contig of 820 bp in length

```

```

* 53822 53921: gap of 100 bp
53922 55355: contig of 1434 bp in length
55356 55455: gap of 100 bp
55456 57801: contig of 2346 bp in length
57802 57901: gap of 100 bp
57902 59549: contig of 1648 bp in length
59550 59649: gap of 100 bp
59650 60555: contig of 906 bp in length
60556 60655: gap of 100 bp
60656 62781: contig of 2126 bp in length
62782 62881: gap of 100 bp
62882 64370: contig of 1489 bp in length
64371 64470: gap of 100 bp
64471 65603: contig of 1133 bp in length
65604 65703: gap of 100 bp
65704 67667: contig of 1964 bp in length
67668 67767: gap of 100 bp
67768 69279: contig of 1512 bp in length
69280 69379: gap of 100 bp
69380 71150: contig of 1771 bp in length
71151 71250: gap of 100 bp
71251 73611: contig of 2361 bp in length
73612 73711: gap of 100 bp
73712 75366: contig of 1655 bp in length
75367 75466: gap of 100 bp
75467 78279: contig of 2813 bp in length
78280 78379: gap of 100 bp
78380 80372: contig of 1993 bp in length
80373 80472: gap of 100 bp
80473 82236: contig of 1764 bp in length
82237 82336: gap of 100 bp
82337 84113: contig of 1777 bp in length
84114 84213: gap of 100 bp
84214 86315: contig of 2102 bp in length
86316 86415: gap of 100 bp
86416 89969: contig of 3554 bp in length
89970 90069: gap of 100 bp
90070 92420: contig of 2351 bp in length
92421 92520: gap of 100 bp
92521 95426: contig of 2906 bp in length
95427 95526: gap of 100 bp
95527 97256: contig of 1730 bp in length
97257 97356: gap of 100 bp
97357 100888: contig of 3532 bp in length
100889 100988: gap of 100 bp
100989 103643: contig of 2655 bp in length
103644 103743: gap of 100 bp
103744 106545: contig of 2802 bp in length
106546 106645: gap of 100 bp
106646 111055: contig of 4410 bp in length
111056 111155: gap of 100 bp

Query Match      20.1%; Score 532.8; DB 2; Length 149830;
Best Local Similarity 99.6%; Pred. No. 1.9e-95;
Matches 534; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1  tgtctagccatagctggcagtcaccgagcgaggctctcagcctgcagagacagacag 60
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 38029 TGTCTAGGCCATAGCTTGGCAGTCCCGGGGGGGGGCTCTCAGCCTGGCAGAGAGCGAG 38088

Qy 61  gacctcacgggggaaaggagctggacgcgcctggcgcgagtggtgggctggcagcgagg 120
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 38089 GACCTTTACGGGGGAAAGGGGCTGGACGCGCTGGCCGCGGTGGGGCTGGCAGCGGGG 38148

Qy 121  cggagaaagcggcgatgccccgggggcttggggatgggagtcacaggggggctcccg 180
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 38149 CGGAAGGAAAGCGCGGATGCCCGGGGGGCTTTTGGGATGGGCGATCCAGGGGGGTCCCG 38208

Qy 181  gagaggggagcagacagaccagagctgtagggagctggaagaccccccagctctact 240
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 38209 GAGAGGGGGAGCAGACAGACCGGAAGGCTGTGTGAGGGGGGGGGAACCCGAGGCTCTCT 38268

Qy 241  gcagggcaagggtccttgctgtagcggggggagcgcgctcttgcctcccggggtcgtgc 300

```

```
|||||
Db 38269 GCAGGCAAGGTCCTGTCTGTACGGGGGACGGCGCTCTTGTCCCGGGGTCTGTGC 38328
Qy 301 agactacgccccctactgccccccacttctctgaccaggggtgcccattctgagtcgcc 360
|||||
Db 38329 AGACTACGGGGCCCTACTGCCCCCCTTCTCTGGACCGGGGTGCCATCTGAGTCCC 38388
Qy 361 tgggggaggggccccctcgggtttttacagacgcccccgctccgctggccagtgctgcc 420
|||||
Db 38389 TGGGGGAGGGGGCCCCCTCGGCGCTTTGACGACGCGCGCTCCGCTGGCGAGGTCGTCCA 38448
Qy 421 tcgcacagcgcacagcatccaccagcgggtcgagagagcctcagcagacacctatc 480
|||||
Db 38449 TCCGCAACGGCGCACAGCATCATCAGCGGTGCGGAGAGCGCCTCAGCGAGACGCCATC 38508
Qy 481 gcccaacagcgtcgtcttctgtcagcagcagcgcgcccccgctccgctggccagtgcc 536
|||||
Db 38509 GCCCAACAGGTCGCTTCTGTCGACGAGCGACGCCCGACGGCGCGCGGACCG 38564

RESULT 13
AC027219/c
LOCUS
DEFINITION Homo sapiens clone RP11-729P19, *** SEQUENCING IN PROGRESS ***, 79
unordered pieces.
AC027219
VERSION AC027219.3 GI:13184220
KEYWORDS HTG; HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 149830)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens, clone RP11-729P19
Unpublished
2 (bases 1 to 149830)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Bida,F.,
Boguslavskiy,L., Boukhalter,B., Brown,A., Burkett,G.,
Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., DeArelano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
Gadagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., LaRoque,K., Lamazares,R., Landers,T., Lehoczy,J.,
Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
McLirim,J., Meneus,L., Mihova,T., Miranda,C., Miengua,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neil,D., Ollivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (28-MAR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 2, 2001 this sequence version replaced gi:11610941.
All repeats were identified using RepeatMasker:
Snit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L8346
Center clone name: 729_P_19
```

* NOTE: This is a 'working draft' sequence. It currently
* consists of 79 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1569: contig of 1569 bp in length
* 1570 1669: gap of 100 bp
* 1670 2364: contig of 695 bp in length
* 2365 2464: gap of 100 bp
* 2465 3099: contig of 635 bp in length
* 3100 3199: gap of 100 bp
* 3200 3762: contig of 563 bp in length
* 3763 3862: gap of 100 bp
* 3863 4534: contig of 672 bp in length
* 4535 4634: gap of 100 bp
* 4635 5574: contig of 840 bp in length
* 5575 5747: gap of 100 bp
* 5748 6560: contig of 886 bp in length
* 6561 7785: contig of 1225 bp in length
* 7786 7885: gap of 100 bp
* 7886 8835: contig of 850 bp in length
* 8836 8864: gap of 100 bp
* 8865 9964: contig of 1029 bp in length
* 9965 10784: contig of 820 bp in length
* 10785 10884: gap of 100 bp
* 10885 11653: contig of 769 bp in length
* 11654 11753: gap of 100 bp
* 11754 12415: contig of 662 bp in length
* 12416 12515: gap of 100 bp
* 12516 13232: contig of 717 bp in length
* 13233 13332: gap of 100 bp
* 13333 14178: contig of 846 bp in length
* 14179 14278: gap of 100 bp
* 14279 15762: contig of 1484 bp in length
* 15763 15862: gap of 100 bp
* 15863 16423: contig of 561 bp in length
* 16424 16523: gap of 100 bp
* 16524 17467: contig of 944 bp in length
* 17468 17567: gap of 100 bp
* 17568 18533: contig of 966 bp in length
* 18534 18633: gap of 100 bp
* 18634 19882: contig of 1249 bp in length
* 19883 19982: gap of 100 bp
* 19983 21070: contig of 1088 bp in length
* 21071 21170: gap of 100 bp
* 21171 22152: contig of 982 bp in length
* 22153 22252: gap of 100 bp
* 22253 23303: contig of 1051 bp in length
* 23304 23403: gap of 100 bp
* 23404 24381: contig of 978 bp in length
* 24382 24481: gap of 100 bp
* 24482 26183: contig of 1702 bp in length
* 26184 26283: gap of 100 bp
* 26284 27663: contig of 1380 bp in length
* 27664 27763: gap of 100 bp
* 27764 28566: contig of 803 bp in length
* 28567 28666: gap of 100 bp
* 28667 30309: contig of 1643 bp in length
* 30310 30409: gap of 100 bp
* 30410 31434: contig of 1025 bp in length
* 31435 31534: gap of 100 bp
* 31535 32225: contig of 691 bp in length
* 32226 32325: gap of 100 bp
* 32326 33690: contig of 1365 bp in length
* 33691 33790: gap of 100 bp
* 33791 34434: contig of 644 bp in length
* 34435 34534: gap of 100 bp

McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R.,
Meldrum, J., Meneus, L., Mihova, T., Miranda, C., O'Connor, T., O'Donnell, P.,
Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
O'Neil, D., Oliver, T.M., Oliver, J., Peterson, K., Pierre, N.,
Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Tesfaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J.,
Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,
Young, G., Zainoun, J., Zimmer, A. and Zody, M.

TITLE

JOURNAL

COMMENT

Submitted (30-APR-2000) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www.seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L9566

Center clone name: 225_K_6

* NOTE: This record contains 74 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

1
* 664 763: gap of 100 bp
* 764 1454: contig of 691 bp in length
* 1455 1554: gap of 100 bp
* 1555 2227: contig of 673 bp in length
* 2228 2327: gap of 100 bp
* 2328 2833: contig of 506 bp in length
* 2834 2933: gap of 100 bp
* 2934 3614: contig of 681 bp in length
* 3615 3714: gap of 100 bp
* 3715 4404: contig of 690 bp in length
* 4405 4504: gap of 100 bp
* 4505 5193: contig of 689 bp in length
* 5194 5293: gap of 100 bp
* 5294 5959: contig of 666 bp in length
* 5960 6059: gap of 100 bp
* 6060 6717: contig of 658 bp in length
* 6718 6817: gap of 100 bp
* 6818 7486: contig of 669 bp in length
* 7487 7586: gap of 100 bp
* 7587 8278: contig of 692 bp in length
* 8279 8378: gap of 100 bp
* 8379 9040: contig of 662 bp in length
* 9041 9140: gap of 100 bp
* 9141 9808: contig of 668 bp in length
* 9809 9908: gap of 100 bp
* 9909 10573: contig of 665 bp in length
* 10574 10673: gap of 100 bp
* 10674 11375: contig of 702 bp in length
* 11376 11475: gap of 100 bp
* 11476 12157: contig of 682 bp in length
* 12158 12257: gap of 100 bp
* 12258 12954: contig of 697 bp in length
* 12955 13054: gap of 100 bp
* 13055 13746: contig of 692 bp in length
* 13747 13846: gap of 100 bp
* 13847 14524: contig of 678 bp in length
* 14525 14624: gap of 100 bp
* 14625 15267: contig of 643 bp in length
* 15268 15367: gap of 100 bp
* 15368 16052: contig of 685 bp in length
* 16053 16152: gap of 100 bp
* 16153 16832: contig of 680 bp in length
* 16833 16932: gap of 100 bp
* 16933 17596: contig of 664 bp in length
* 17597 17696: gap of 100 bp
* 17697 18373: contig of 677 bp in length
* 18374 18473: gap of 100 bp
* 18474 19145: contig of 672 bp in length
* 19146 19245: gap of 100 bp
* 19246 19917: contig of 672 bp in length
* 19918 20017: gap of 100 bp
* 20018 20705: contig of 688 bp in length
* 20706 20805: gap of 100 bp
* 20806 21500: contig of 695 bp in length
* 21501 21600: gap of 100 bp
* 21601 22261: contig of 661 bp in length
* 22262 22361: gap of 100 bp
* 22362 23019: contig of 658 bp in length
* 23020 23119: gap of 100 bp
* 23120 23786: contig of 667 bp in length
* 23787 23886: gap of 100 bp
* 23887 24571: contig of 685 bp in length
* 24572 24671: gap of 100 bp
* 24672 25338: contig of 667 bp in length
* 25339 25438: gap of 100 bp
* 25439 26102: contig of 664 bp in length
* 26103 26202: gap of 100 bp
* 26203 26863: contig of 661 bp in length
* 26864 26963: gap of 100 bp
* 26964 27643: contig of 680 bp in length
* 27644 27743: gap of 100 bp
* 27744 28444: contig of 701 bp in length
* 28445 28544: gap of 100 bp
* 28545 29219: contig of 675 bp in length
* 29220 29319: gap of 100 bp
* 29320 30005: contig of 686 bp in length
* 30006 30105: gap of 100 bp
* 30106 30783: contig of 684 bp in length
* 30790 30889: gap of 100 bp
* 30890 31572: contig of 683 bp in length
* 31573 31672: gap of 100 bp
* 31673 32347: contig of 675 bp in length
* 32348 32447: gap of 100 bp
* 32448 33134: contig of 687 bp in length
* 33135 33234: gap of 100 bp
* 33235 33916: contig of 682 bp in length
* 33917 34016: gap of 100 bp
* 34017 34701: contig of 685 bp in length
* 34702 34801: gap of 100 bp
* 34802 35480: contig of 679 bp in length
* 35481 35580: gap of 100 bp
* 35581 36273: contig of 693 bp in length
* 36274 36373: gap of 100 bp
* 36374 37033: contig of 660 bp in length
* 37034 37133: gap of 100 bp
* 37134 37795: contig of 662 bp in length
* 37796 37895: gap of 100 bp
* 37896 38553: contig of 658 bp in length
* 38554 38653: gap of 100 bp
* 38654 39267: contig of 614 bp in length
* 39268 39367: gap of 100 bp
* 39368 40053: contig of 686 bp in length
* 40054 40153: gap of 100 bp
* 40154 40842: contig of 689 bp in length
* 40843 40942: gap of 100 bp
* 40943 41629: contig of 687 bp in length
* 41630 41729: gap of 100 bp
* 41730 42416: contig of 687 bp in length
* 42417 42516: gap of 100 bp


```
/gene="NCF1"
/notes="polymorphism; compared to sequence deposited in
GenBank Accession Number AF330626"
/replace="g"

BASE COUNT      288 a      441 c      394 g      217 t
ORIGIN

Query Match      10.5%; Score 279; DB 9; Length 1340;
Best Local Similarity 92.2%; Pred. No. 3e-45;
Matches 294; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

QY 851 gcggggcggggctgtgactgcccgcgtctctgcccgcagaggagcgagcgagcgc 910
    || |||| |||| || || || || || || || || || || || || || || || ||
Db 1020 GCCAGGCGCGCGGACCGCGAGAGCCCGGAGCGCGCTCGAGGAGGAGCGGACGCGC 1079

QY 911 agcgcttaaacgcagcgcggtgccccgcggcgagcgccgacctcctgaacc 970
    || || || || || || || || || || || || || || || || || || || ||
Db 1080 AGCGCTCTAAACCGCAGCCGCGGTGCCCGCGCGGCGAGCGCGACCTCATCCTGAACC 1139

QY 971 gctgcagcgagagcaccgaagctggtgctctgctgctgaggtgagcgagcagtc 1030
    || || || || || || || || || || || || || || || || || || || ||
Db 1140 GCTGCACGAGAGACCAACAGCGGAGCTGGCGTCTGCCGTCTGAGGCTGAGCGCAGTCC 1199

QY 1031 ccagctagcgtctgcgccttgccgccttgctgtatatacgtgttctatagagcctg 1090
    || || || || || || || || || || || || || || || || || || || ||
Db 1200 CCAGCTAGCGTCTCGGCGCTTGCGCGCCCTGCCTGTACATACGTGTCTATAGAGCCTG 1259

QY 1091 gcgtctgagcgagggagggagcccgacctgtccagcgcgctcccgccaccctcaat 1150
    || || || || || || || || || || || || || || || || || || || ||
Db 1260 GCGTCTGGACGCCGAGGGGACGCCCGACCCCTGTCTCCAGCGCGGCTCCCGCACCCCTCAAT 1319

QY 1151 aaatgttgcctggagtggg 1169
    || || || || || || || || || || || || || || || || || || || ||
Db 1320 AAATGTTGCTTGGAGTGGGA 1338
```

Search completed: January 11, 2002, 18:02:26
Job time: 32098 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 11, 2002, 09:07:28 ; Search time 8779.43 Seconds
(without alignments)
1691.530 Million cell updates/sec

Title: US-09-820-005-1
Perfect score: 1382
Sequence: 1 cctggaagtgcaggagca.....aaaaaaaaaaaaaaaaaaaaa 1382

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST.*

- 1: em_estfun.*
- 2: em_esthum.*
- 3: em_estlin.*
- 4: em_estom.*
- 5: em_estpl.*
- 6: em_estba.*
- 7: em_estro.*
- 8: em_estov.*
- 9: em_hic.*
- 10: gb_est1.*
- 11: gb_est2.*
- 12: gb_hic.*
- 13: gb_gss.*
- 14: em_gss_fun.*
- 15: em_gss_hum.*
- 16: em_gss_inv.*
- 17: em_gss_pln.*
- 18: em_gss_pro.*
- 19: em_gss_rod.*
- 20: em_gss_vrt.*
- 21: em_gss_other.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	914.6	66.2	976	10	AL582825
C 2	890.2	64.4	1029	10	AL577485
C 3	867	62.7	938	10	AL554867
C 4	843.4	61.0	952	10	AL550056
C 5	836.2	60.5	914	10	AL574020
C 6	833.6	60.3	904	10	AL583217
C 7	824.6	59.7	938	10	AL559806
C 8	813.8	58.9	895	10	AL583422
C 9	800.6	57.9	906	10	AL561079
C 10	789.8	57.1	854	10	AL560372
C 11	787.8	57.0	889	10	AL583335
C 12	775.6	56.1	954	10	AL561678

13	723.4	52.3	851	11	BG758136
14	719.2	52.0	1070	11	BF974789
C 15	712.2	51.5	834	10	AL581347
C 16	703.2	50.9	978	10	BE561619
C 17	701	50.7	1042	11	BF976162
C 18	700.4	50.7	1108	11	BG106156
C 19	699.4	50.6	735	11	BG696253
C 20	696.4	50.4	1015	11	BG755063
C 21	688.4	49.8	752	11	BF976456
C 22	683.4	49.5	977	11	BF97068
C 23	682.2	49.4	751	11	BG398248
C 24	669.6	48.5	745	11	BG756837
C 25	640.2	46.3	870	10	BE561667
C 26	635.4	46.0	846	10	AL582483
C 27	634.2	45.9	707	10	AL583276
C 28	609.6	44.1	698	10	AW340740
C 29	598.6	43.3	887	11	BF129021
C 30	593.2	42.9	660	10	BE268407
C 31	591	42.8	617	10	BE397098
C 32	586.4	42.4	813	11	BF128894
C 33	583.4	42.2	656	11	BI227357
C 34	582.2	42.1	750	10	AW072388
C 35	570.6	41.3	602	11	BI224931
C 36	560.8	40.6	877	10	AA767386
C 37	553.6	40.1	608	10	BE397270
C 38	542.4	39.2	661	10	AA742367
C 39	541.4	39.2	551	10	BE501539
C 40	536	38.8	566	11	BF064049
C 41	533.8	38.6	547	10	AI436579
C 42	526	38.1	535	10	AI439666
C 43	524.8	38.0	535	11	BE855380
C 44	524.8	38.0	625	10	AI479089
C 45	520	37.6	545	10	BE513934

ALIGNMENTS

RESULT 1	AL582825/c	976 bp	mRNA	EST	16-FEB-2001
LOCUS	AL582825	LT1_NFL010_BC2	Homo sapiens	cDNA clone	CS0DL008YK23 3
DEFINITION	AL582825	prime, mRNA	sequence.		
ACCESSION	AL582825				
VERSION	AL582825.1	GI:12951193			
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	1 (bases 1 to 976)				
AUTHORS	Li, W.B., Gruber, C., Jessee, J., and Polayes, D.				
TITLE	Full-length cDNA libraries and normalization				
JOURNAL	Unpublished (2001)				
COMMENT	Contact: Genoscope Genoscope - Centre National de Sequencage Bp 191 91006 Evry cedex - France Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.				
FEATURES	Location/Qualifiers				
source	1. .976				
	/organism="Homo sapiens"				
	/db_xref="taxon:9606"				
	/clone="CS0DL008YK23"				
	/clone_lib="LT1_NFL010_BC2"				
	/sex="male"				
	/tissue_type="B cells from Burkitt lymphoma"				
	/note="Vector: pCMVSPORT 6; Site 1: NotI; last strand cDNA was primed with a NotI-oligo(GT) primer. Five prime end enriched, double-stranded cDNA was digested with NotI and cloned into the NotI and EcoRV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive				

Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com

BASE COUNT 147 a 293 c 326 g 205 t 5 others
ORIGIN

```
Query Match 66.2%; Score 914.6; DB 10; Length 976;
Best Local Similarity 97.2%; Pred. No. 6.5e-128;
Matches 950; Conservative 5; Mismatches 9; Indels 13; Gaps 2;

Qy 335 acgctcatgagcctgccaccacaaagatctccgcgtgtcccccacctctcgtactcttcaag 394
Db 976 AGCCTCATGAGCCIGCCACCAAGATCTCCGCTGTCCCACTCTCTTCTTCAAG 917

Qy 395 gtggccctgtatgaactcaagctcccccagacacacacacacacacacacacacacacac 454
Db 916 GTGGCCCTGTATGACTCAAGCTCCCAAGTCCCAAGACACACACACACACACACACACATAC 857

Qy 455 ttgatgcccaagatggcaagatgacacacacacacacacacacacacacacacacacacac 514
Db 856 TTGATRCAAAARATGGCAAGATGACCCGACACACATCACCGGCCCATCATCTCTGCAG 797

Qy 515 acgtaccgcgcattgccaactcagagagacacacacacacacacacacacacacacacac 574
Db 796 AGTACCGCGCATTTGCCGACTACGAGAAGACCTCGGCTCCGAGATGGCTCTGTCCACG 737

Qy 575 ggggaagctgggtggaggtctgtgagagagagagagagagagagagagagagagagagag 634
Db 736 GGGGACGTGGTGGAGGTCTGGAGAAGACGAGAGCGGTTTGGTGTCTGTGCAGATGAAA 677

Qy 635 gcaagcagagctgagatccacagctctctcagagcctcggagcctcggagcctcggagcag 694
Db 676 GCAAGCAGAGCTGGATCCACAGATCTCTCTCGAGCCCTTGGACAGCTCTCTGACAGACG 617

Qy 695 gaagaccctgagccaaactatgcaggtgagccatacgtcgccatcaagggcctacactgct 754
Db 616 GAAGACCCTGAGCCCACTATGACAGTGAGGTCAGGCATACGTCGCCATCAAGGCCCTACACTGCT 557

Qy 755 gtgagagggagcagaggtgtccctcgtcagaggtgaagctgtgaggtcattcacaagctc 814
Db 556 GTGAGGGGGACGAGGTGTCCCTGCTCGAGGGTGAAGCTGTGAGGTGATTCACAAAGCTC 497

Qy 815 ctggagcggct-----ggaagacagcgtccacagcgtacttccgtccatgtac 862
Db 496 CTGGACGGCTGTGGGTGTCATCAGGAAGACCGTCACAGGCTACTTTCCTCCATGTAC 437

Qy 863 ctgcacaaagtcagggaagacggtgtccacagggcccaagccagatcaagcgggggcccgcg 922
Db 436 CTGCAAAAAGTCGGGCAAGAGCTGTCCCAAGGCCCAAGCCAGATCAAGCG-GGGGCGCGC 378

Qy 923 cccgcaggtcgtccatccgcaacgcgcacacagatccacacacagcgtgcggaagcgcctc 982
Db 377 CCCGCAAGTGTCTATCCGCAACGCGCACAGCAPCCATCATCAGCGGTGCGGAAGCGCCTC 318

Qy 983 agcaggagcgcctatcgccgcacacagcgtctcttcttcgacgagcgcgcgcgcgcgcgc 1042
Db 317 AGCCAGGAGCCTATCGCCGCAACAGAGTGTCTTCTGACAGACGACGCGCCGACGCGC 258

Qy 1043 cggccggacccgagagcccccggagcccgctcagagagagagcgcgcgcgcgcgcgcgc 1102
Db 257 CGGCGGGACCGCAGAGCCCGGGAGCCCGCTCGAGGAGGAGCGCAGACGACGCGCTCT 198

Qy 1103 aaacccagcggcgggtgtcccccggcggcggcggcggcggcggcggcggcggcggcggc 1162
Db 197 AAACCGAGCGGGGGTGTCCCGGGCGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 138

Qy 1163 gagagcaccacggaagctgtgctgtcgcgtctgaggtcgtgagcgcagtcgagtcgagtcag 1222
Db 137 GAGAGCACCAAGCGGAAGCTGGCGTCTGCGCTCTGAGGCTGGAGCGCAGTCCCCAGCTAG 78

Qy 1223 cgtctcggcccttgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1282
Db 1282 cgtctcggcccttgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc
```

Db 77 CGTCTGGCCCTTGGCGCCCGTGCCTGTACATAGCTGTCTTATAGAKCCTGGCGTCTGG 18
Qy 1283 acgcgagagggcagcccc 1299
Db 17 ACGCCGAGGCGAGCCCC 1

RESULT 2
AL577485/c

LOCUS AL577485 1029 bp mRNA EST 16-FEB-2001
DEFINITION AL577485 LTI_NFL006.PL2 Homo sapiens cDNA clone CS0DI087YH09 3
prime, mRNA sequence.

ACCESSION AL577485

VERSION AL577485.1 GI:12940661

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 1029)
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS Li, W.B., Gruber, C., Jessee, J., and Polayes, D.

TITLE Full-length cDNA libraries and normalization

JOURNAL Unpublished (2001)

COMMENT Contact: Genoscope

Genoscope - Centre National de Sequencage

BP 191 91006 EVRY cedex - France

Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

Location/Qualifiers

1. 1029

source

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="CS0DI087YH09"

/clone_lib="LTI_NFL006.PL2"

/tissue_type="placenta"

/note="vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA

was primed with a NotI-oligo(dT) primer. Five prime end

enriched, double-stranded cDNA was digested with Not I and

cloned into the Not I and Eco RV sites of the pCMVSPORT 6

vector. Library was normalized. Library was constructed by

Life Technologies. Contact : Feng Liang Life Technologies,

a division of Invitrogen 9800 Medical Center drive

Rockville, Maryland 20850, USA Fax : (1) 301 610 8371

Email : fliang@lifetech.com URL :

http://fulllength.invitrogen.com"

BASE COUNT 148 a 293 c 359 g 209 t 20 others

ORIGIN

Query Match 64.4%; Score 890.2; DB 10; Length 1029;

Best Local Similarity 93.4%; Pred. No. 2.8e-124;

Matches 961; Conservative 8; Mismatches 43; Indels 17; Gaps 4;

Qy 310 gggcacactaccagtagtactgcagcagctcatgagcgtgccccacca-agaatctccgcgt 368

Db 1029 GGGMAAACTTACCAGAACTCGGCACGCTCATGAGCGTGCACCCAGTCTCCCGCT 970

Qy 369 gtccccactctcgtactcttcaaggtgcgcctctgagcctcaagctcccccagaca 428

Db 969 GTCCCCACCTCTCCACTTCTTCAGGTGCGCCCTGTATGACTCAAGCTCCCCACGGACA 910

Qy 429 accagac 488

Db 909 ACCAGACAAAAAGCCAGAGACACACACACACACACACACACACACACACACACACAC 850

Qy 489 acatcacccgccccatcatctcgtcagacgtaccgcgcattgccaactacagagaacct 548

Db 849 ACATMACCGGCCCATCATCTCGGACAGCTACCGCGCATTTGCCAAGTACAGAGAGACT 790

Qy 549 cgggctccagagatgctcttccacgggggagcgtggtgaggtcgtagagagagacagaga 608

Db 789 CGGCTCCGAGATGGCTCTGTCCACGGGGACGTTGGTGGAGGTCTGTAGAGAGACCGAGA 730

Qy 609 gcggtgtgtgtctgtcagatgaagcgaagcgaagcgtggtggtatccacagcgtcttctctcg 668

```
|||||
Db 729 GCGGTTGGTGTCTGTGATGAAGCAAGCGAGGCTGGATCCCAAGCGTCTCTCTCG 670
Qy 669 agccccggacagtccttgacagaggaagacccctgagcccaactatgcaggtagacccat 728
Db 669 AGCCCCGTCACAGTCTTGACGANACGGAAGACCCCTGAGCCCACTATGCAGNTGAGCCAT 610
Qy 729 acgtcgccatcaagggctacactgctgtgagggggagcaggtgtcctctgtcgaggggtg 788
Db 609 ACGTCGCCATCAAGGCTTANACTGCTGTGAGGGGAGCAGAGGTCTCCTGCTGTCGAGGGTG 550
Qy 789 aagctgttgaggtcatcaagagctctgacggct-----gaaagacagc 836
Db 549 AAGCTGTGAGGTWATTNACAAAGTCTCTGACGGCTGGTGGTCCCTCAGGAAGACGACG 490
Qy 837 tcacagctacttccctccatgtacctgcaaaagtgcaggaagcgtgtccccagccc 896
Db 489 TCACAGGCTACTTCCCGTCCATGTACTGTCNAAAGTCAAGGCNAGAGCTGTCCAGGCC 430
Qy 897 aacccagatcaagggggggcgcccgcccgaggtgtccatccgcaacgcgcacagca 956
Db 429 AACGCCAGTCAAGCGGGGGCGCGCCCGCAGGTCGTCCATCCGNCNACGNCAGCA 370
Qy 957 tccaccagggctgcggaagcgcctcagcagagcgcctatcgccgcaaacagcgtccgtt 1016
Db 369 TCCACCAAGCGTCCGGAAGCGCTCAGCCAGNACGGCTATCGCGCGNACAGCGTCCGTT 310
Qy 1017 tctgcagcagcagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1076
Db 309 TTTGTCAGCAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 250
Qy 1077 agggagcgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1135
Db 249 AGGAGGAGTGCACAGCAGCGCTCTTAACCGCAGCGCGCGCGCGCGCGCGCGCGCGAG 190
Qy 1136 gccagctctctgaaccgctgcagcagagcagcagcagcagcagcagcagcagcagcagc 1195
Db 189 GCCGACCTCATCTTGAACCGCTGAGCGAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 130
Qy 1196 tgagcttgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1252
Db 129 TGAGCTGGAGCGGAGTGCACACAGCTAGCATCTCGCCCTTCCGCGCGCGCGCGCGCTG 70
Qy 1253 taccgtgttctatagagcctggcgtctggagcgcggagcgcggagcgcggagcgcggagc 1312
Db 69 CATAGCTGTCTATAGAGCTGGGCTGTGGAGCGCCAGCGCCCGCGCGCGCGCGCGCGCC 10
Qy 1313 cgcggctcc 1321
Db 9 CCCCCCCCC 1

RESULT 3
AL554867 938 bp mRNA EST 16-FEB-2001
LOCUS AL554867 LTI_NFL006.PL2 Homo sapiens cDNA clone CS0DI087YH09 5
DEFINITION prime, mRNA sequence.
ACCESSION AL554867
VERSION AL554867.1 GI:12896059
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 938)
AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
Location/Qualifiers

FEATURES
```

```
source 1..938
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DI087YH09"
/tissue_type="placenta"
/notes="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com"
BASE COUNT 222 a 288 c 262 g 159 t 7 others
ORIGIN

Query Match 62.7%; Score 867; DB 10; Length 938;
Best Local Similarity 96.9%; Pred. No. 8.6e-121;
Matches 913; Conservative 6; Mismatches 7; Indels 16; Gaps 3;

Qy 43 ggacacatttcacgtccacatcgccctgctgggtttgagagcgtctctaccagcca 102
Db 1 GCACACCTTTCATCGCTCACATCGCCCTGCTGGGCTTTGAGAAGCGCTTCGTACCGACGCA 60
Qy 103 gcactatgtatcatgttctctggtgaaatggcagagcgtgtcgagaaggtggtctaccg 162
Db 61 GCATATGTGTACATGTTCTCTGGTGAATGGCAGACCTGTCCGAGAAGGTTGGTCTACCG 120
Qy 163 gcgcttcacagagatctacgagttccataaaacaccttaaaagaaatgttccctattgagc 222
Db 121 GCGCTTCAACGAGATCTACGAGTWCCTATAAACCTTAAAGCAANTGTTCCTATTGAGGC 180
Qy 223 agggcgatcaatccagagaacaggtatcatccccacactccagctcccaagtggttga 282
Db 181 AGGGCGATCAATCAGAGAACAGGATCATCCCCACCTCCAGCTCCCAAGTGTGTA 240
Qy 283 cggcagcggggcccgagaaacccaggggcaacttaccgagtagtactgcagcagcctcat 342
Db 241 CGGCAGCGGGCCCGCAGAAACMGCCAGGACACATTACCAGTACTGCACAGCGCTCAT 300
Qy 343 gagctgcaccaagatctcccgctgtccccacactcctcagctcttcaaggtagcgc 402
Db 301 GAGCWTGCCACCAAGATCTCCCGCTGTCCCGACCTCTCTCAGCTTCTTCAAGTGGCGCC 360
Qy 403 ttagctcactcagctcccccacggacacacagcaaaaaagccagacatacttgatgcc 462
Db 361 TGATGACCTCAAGCTCCCGACGACACACAGACAAAAAAGCCAGAGACATACTTGATGCM 420
Qy 463 caaagatggcaagagtagccgcagacacatcacggccccccatcctcgacagagtagccg 522
Db 421 CAAGATGGCAAGAGTAGCCGCGCAGACATACCCGGCCCCCATCTCCGTCAGAGCATACCG 480
Qy 523 cgcatatgcacactacagaaagcgtcggtccgagatggctctgtccacggggagcgt 582
Db 481 CGCATTTGCCAATCAGCAGAGACCTCGGGCTCGCAGATGGCTCTGTCCACGGGGAGCT 540
Qy 583 ggtggaggtctagagaagcgcagagcgtgtggtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 642
Db 541 GGTGGAGGTCTAGAGAAGAGCAGAGCGGNTTGGTGGTCTCTGTGACAGTGAAGCAAGACG 600
Qy 643 aggtggtacccagcgtctctctcgagccccctggagagctcctcgagagaggaagacc 702
Db 601 AGGTGGATCCCGCGCTCTTCTCGAGCCCCCTGGACAGTCTCTGACGAGACGGAAGACCC 660
Qy 703 tgaccacactatcaggtgagccatcgtcgccatcagcctacactgctgtgtgtgtgtgtgt 762
Db 661 TGACCCCAACTATGCAAGTGAGCCATACGTGCCCATCAAGGCCCTACACTGCTGTGGAGGG 720
Qy 763 ggaagaggtgtccctgctcgaggggtgaagcgtgtgaggtcatttcacaagcctctgagcgg 822
```

Db 721 GGACGAGGTGTCCTGCTCGAGGGTGAAGCTTTGAGGTTCATTCACAGCTCCTCGACGG 780
 Qy 823 ct-----ggaagacagcgtcacaggctacttccgctccatgtacctgcaaaa 870
 Db 781 GTGGTGGGTTCATCAGGAAGACGAGCTCACAGGCTACTTCGGCTCCATGTACTGCAAAA 840
 Qy 871 gtccagggcaagcgtgtccagggcccaagccagacatcaagcggggcgccgcgccgcag 930
 Db 841 GTACGGCAAGACGTGTCCAGGCCACAGCCAGCAAGCGGGGGGGCGC--GCCCGCAG 898
 Qy 931 gtccgtccatccgcgaacgcgcagcaccatccaccagcggctgcg 972
 Db 899 GTCGTCCATCCGCAA--CGGCACASATCCACACGCGGTGCGG 938

RESULT 4

AL550056 952 bp mRNA EST 16-FEB-2001
 LOCUS AL550056 LTI_NFL006_PL2 Homo sapiens cDNA clone CS0DI053YF12 5
 DEFINITION prime, mRNA sequence.

ACCESSION AL550056

VERSION AL550056.1 GI:12886644

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS Li, W.B., Gruber, C., Jessee, J., and Polayes, D.

TITLE Full-length cDNA libraries and normalization

JOURNAL Unpublished (2001)

COMMENT Contact: Genoscope

Genoscope - Centre National de Sequencage

BP 191 91006 Evry cedex - France

Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

Location/Qualifiers

1. 952

source

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="CS0DI053YF12"

/clone_lib="LTI_NFL006_PL2"

/tissue_type="placenta"

/note="Vector: pCMVSPORT 6; Site.1: NotI; 1st strand cDNA

was primed with a NotI-oligo(dT) primer. Five prime end

cloned into the Not I and Eco RV sites of the pCMVSPORT 6

vector. Library was normalized. Library was constructed by

Life Technologies. Contact : Feng Liang Life Technologies,

a division of Invitrogen 9800 Medical Center Drive

Rockville, Maryland 20850, USA Fax : (1) 301 610 8371

Email : fliang@lifetech.com URL :

http://fulllength.invitrogen.com"

BASE COUNT 223 a 299 c 263 g 163 t 4 others

ORIGIN

Query Match 61.08; Score 843.4; DB 10; Length 952;

Best Local Similarity 95.18; Pred. No. 2.9e-117;

Matches 931; Conservative 2; Mismatches 15; Indels 31; Gaps 5;

Qy 48 ccttcacgtcacatgcctcgtggcctttgagagcgtctcgtaccagccagcact 107

Db 1 CCTTCATCCGTACATCCCTCGTGGCTTTGAGAGCGCTTCGTACCCAGCCAGCACT 60

Qy 108 atgtgtacatgttctcgtgtaaaaggcaggacctcgcgagaaggtggtctaccgcgcgt 167

Db 61 A--TGTACATGTTCTGTGTAATGGCAGGACCTGTCCGAGAAAGTGGTCTACCGGCGCT 118

Qy 168 tcaccagatctacagttccataaaaccttaaaagaatttccctattgaggcgggg 227

Db 119 TCACCGAGATCTAGAGTTCATAAAAGCTTTAAAGAAATGTTCCCTATTGAGGCGAGGG 178

Qy 228 cgatcaatccagaaacaggatcatccccacacctccagctcccaagtgggttgacgggc 287
 Db 179 CcATCAATCCAGAGAACAGAGGATCATCCCCCACCCTCCCAAGTGGTGTGACGGGC 238
 Qy 288 agcgggccgcgcagaaacccgcagggcacacttaccaggtacttgagcacgctcatgagcc 347
 Db 239 AGCGGGCCCGCGAGAACCAACAGGGCACACTTACCGAGTACTGCGGCACGCTCATGAGCC 298
 Qy 348 tgcacacaagatctcccgctgtcccccacctctctctctctctctctctctctctctgag 407
 Db 299 TGCCCAACAAGATCTCCCGCTGTCCCACTTCTTCAAGGTGGCGCCCTGATG 358
 Qy 408 acctcaagctccccacagaaacacagacacacacacacacacacacacacacacacac 467
 Db 359 ACCTCAAGCTCCCCACGACCAACAGACAAAAAGCCAGACATACCTGTATGCCCAAG 418
 Qy 468 atggcaagtagtaccgcgcagacacatcacggcccccatctctcgcagacgtaccgcgcga 527
 Db 419 ATGCAACAGTACCGGACAGACATCACCGCCCCCATCATCTCGACAGCTACCGCGCCA 478
 Qy 528 tgcacaactacgagaagacctcgcggctccgagatggctctctcgcacggggagcgtggtg 587
 Db 479 TTGCCAACTACGAGAAGACCTCGGGCTCCGAGATGGCTCTGCCACGCGGGACGCTGTGG 538
 Qy 588 aggtcgttagaagagcagagcgggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 647
 Db 539 AGGTGCTGGAGAGCGGAGCGGNTTGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 598
 Qy 648 ggaatccacgctctctcgcagccccctggagctcctcgcagcagcagcagcagcagcagc 707
 Db 599 GGATCCCAAGCATCTTCTCGAGCCCCCTGGACAGTCTCTGACGAGCGGACGACACCTGAGC 658
 Qy 708 ccaactatgagggtgagccatcagtcgccatcaaggccacactgctgtgagggggagcgt 767
 Db 659 CCAACTATTCAGAGTGAGCCATACGTGCGCATCAAGGCGCTACACTGCTGTGGAGGGGAGC 718
 Qy 768 aggtgtccctcgtcagaggtgaaagctgttgaggtcattcacaagctcctcgcagcgtgga 827
 Db 719 AGGTGTCTCTGCTCGAGGGTGAAGCTGTGA-----GGG 752
 Qy 828 aagcagcgtcacaggctacttcccgtccatgtacctgcacaaagtccaggcgaagcagcgtgt 887
 Db 753 AAGACGAGCTCATAGGCTACTTCCGCTCATGTACTTCTGCAAAAGTACGAGCGACGCTGT 812
 Qy 888 cccagggcccaacgcagatcaagcgggggggggggggggggggggggggggggggggggggg 947
 Db 813 CCCAGGCCCAACGCGAGATCAAGCGGGGGGGG--CGCCGCGCAGGTGCGTCCATCGGCAACG 871
 Qy 948 cgcacagcatccaccagcgggtcgcggaagcgcctcgcagcagcagcagcagcagcagcagc 1007
 Db 872 TGCACAGCATCCACAGCGGTGCGGGAAGC--CCTCAGCCAGGACGCGCTATCGCGC--AMA 929
 Qy 1008 gcgtccgtttttctgcagca 1026
 Db 930 GCGTCCGTTTTCTGCAGCA 948

RESULT 5

AL574020/c

LOCUS

DEFINITION AL574020 LTI_NFL006_PL2 Homo sapiens cDNA clone CS0DI053YF12 3

prime, mRNA sequence.

ACCESSION AL574020

VERSION AL574020.1 GI:12933820

KEYWORDS EST.

SOURCE human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 914)

AUTHORS Li, W.B., Gruber, C., Jessee, J., and Polayes, D.

TITLE Full-length cDNA libraries and normalization

JOURNAL Unpublished (2001)

COMMENT	FEATURES	source
<p>Contact: Genoscope Genoscope - Centre National de Sequencage BP 191 91006 EVRY cedex - France Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.</p>	<p>Location/Qualifiers 1. 914</p>	<p>137 a 273 c 305 q 194 t 5 others</p>
	<p>/organism="Homo sapiens" /db_xref="taxon:9606" /clone="CS0DI053YF12" /clone_lib="LTI_NFL006_PL2" /tissue_type="placenta" /notes="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com"</p>	

[illegible]

Qy	1019	ctgcagcagcgacgcccaggcgccgggaccgagcccgagccccggagccgactcgag	1078
Dd	281	CTGCAGCAGCGACGCCGCCAGCGCCGCGGACCGCCGAGCGCCGAGCGCCGCTCGAG	222
Qy	1079	gagagcgcgagcagcgacgctctaaaccgcagcgccgggtgcccccgcgagcgccgcgcgc	1138
Dd	221	GAGGAGCGGACGACGACGACTTAAACCAGCAGCGGGGTGCCCGCGCGCGAGCGCC	162
Qy	1139	gacctatcttgaaccgctgcagcgagcagcaccaagcggaagtgcggcgttcgcgcgtcta	1198
Dd	161	GACCTCATCTTGAAACCGCTGCAGCGAGAGCACCAAGCGGANGCTGGCATCTGCCGTCTGA	102
Qy	1199	ggcttgagcgcgagtcctccagctagcgtctcgcgcctctgcgcgcgcgcgcgttgtataacg	1258
Dd	101	GGCTGGGCGCGAGTCGCCAGCTAGCGCTCTCGGCCCTTTGCCCCCCGCTGCTGTATACG	42
Qy	1259	tgtttctagaagcctgcgctctgacccgagcgacgcc	1298
Dd	41	TGTTCTATAGAGCTGGCGCTTGGACGCCGAGGCGACGCC	2
RESULT 6			
AL583217 904 bp mRNA EST 16-FEB-2001			
LOCUS AL583217 LTI_NFL010_BC2 Homo sapiens cDNA clone CS0DL012YG22 5			
DEFINITION prime, mRNA sequence.			
ACCESSION AL583217			
VERSION AL583217.1 GI:12951967			
KEYWORDS EST.			
SOURCE human.			
ORGANISM Homo sapiens			
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
REFERENCE			
AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.			
TITLE Full-length cDNA libraries and normalization			
JOURNAL Unpublished (2001)			
COMMENT Contact: Genoscope Genoscope - Centre National de Sequencage BP 191 91006 EVRY cedex - France Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr. Location/Qualifiers			
FEATURES			
source			
1..904			
/organism="Homo sapiens"			
/db_xref="taxon:9606"			
/clone="CS0DL012YG22"			
/clone_lib="LTI_NFL010_BC2"			
/sex="male"			
/tissue_type="B cells from Burkitt lymphoma"			
/note="vector: pCMVSPORT 6; Site_1: NotI; first strand cDNA was primed with a NotI-oligo(GT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 5 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com"			
BASE COUNT 213 a 274 c 257 g 158 t 2 others			
ORIGIN			
Query Match 60.3%; Score 833.6; DB 10; Length 904;			
Best Local Similarity 97.1%; Pred. No. 8.6e-116;			
Matches 873; Conservative 1; Mismatches 11; Indels 14; Gaps 60			
Qy	40	gggggacacctcatcgcgtcacatcgccctgctgggctttgagaagcgtctcgaccacg	99
Dd	1	GGGGACACCTTCATCCGTCACATCGCCCTCTGGGCTTTCAGAAGCGCTTCGTACCCAG	60
Qy	100	ccagcactatgtgtacattgttcctggtgaaatggcaggacctgtcgggagaagtggtcta	159

```
Db 61 CCAGCACTA--TGTACATGTTCTCTGGTGAATAGGCAGGACCTGTGCGAGAAGGTGGTCTA 118
Qy 160 cggcgcttcaccagatctacgagttccataaaacaccttaaaaaaagtgtccctattga 219
Db 119 CCGGGCTTACCGAGATCTACGAGTTCATTAACACCTTAAGAAATGTTCCCTATTGA 178
Qy 220 ggcggggcgatcaatccagagagacagatcatccccacacctccagctcccaagtgtt 279
Db 179 GGCAGGGCGATCAATCCAGAGAACAGGATCATCCCCACCTCCAGCTCCCAAGTGGT 238
Qy 280 tgacggcgagggcgccgagagaccgcagggcacacctaccagtagtactgcagacgct 339
Db 239 TGACGGCAGCGGGCGCGGAGAACACACAGGACATCACCGGCCCATCATCTTCGACAGCT 298
Qy 340 catgagcctgccaccaagatctcccgctgtccacacctccacacctccacacctccacacct 399
Db 299 CATGAGCCTGCCACCAAGATCTCCCGTGTCCCGCTCTCCCGCTCTCCCGCTCTCCCGCT 358
Qy 400 ccctgatgacctcaagctcccccaggaacacacagacacacacacacacacacacacacacac 459
Db 359 CCTGTGATCACTCAAGCTCCCGGACACACACAGACACACACACACACACACACACACACAC 418
Qy 460 gcccaaaagatggcaagaatccgcgacagacatcacggcccccatcatcctgcagacgta 519
Db 419 GCCCAAGATGGCAAGATACCGGACAGACATCACCGGCCCATCATCTTCGACAGCTA 478
Qy 520 cgcgcctatcccaactacgagagacctcggctcgcgagatgctgtccacgggga 579
Db 479 CCGGCCAATGCCNACTACGAGAGACCTCGGGCTCCGAGATGCTCTGTCCACGGGGGA 538
Qy 580 cgtggtgaggtcgttagaagaagcgcgagcggttggtggttctgtcagatgaagacaaa 639
Db 539 CGTGTGAGGTCTGTGGAGAAGCGGAGAGCGNTTGTGTGTTCTGTCAGATGAAGCAAA 598
Qy 640 gcgaggttgatccagcgtccttctctcgcgagacctcgcgagacctcgcgagacctcgcgag 699
Db 599 CGGAGGCTGGATCCAGCATCTTCTCGAGCCCTGACAGTCTCTGACGAGACGGAAGA 658
Qy 700 ccctgagcccaactatgcaggtgagcatatcgcccatcaagccctacactgctgtgga 759
Db 659 CCTGAGCCCAACTATGAGGTGAGCATTACCTGCCATCAAGGCCCTACACTGCTGTGGA 718
Qy 760 ggggagcagaggtgctcctgctcagaggtgaaagctgttgagtcattcacaaagctcctgga 819
Db 719 GGGGACAGAGGTGCTCCTGCTCGAGGGTGAAGCTGTTGAGTCAATCACAAGCTCCTGGA 778
Qy 820 cggct-----ggaagacgagctcagaggtcacttcccgctccatgtacctgca 867
Db 779 CGGCTGTGGTCTATCAGGAAGACGACGCTACAGGCTACTTTCCGTCCTATGTACCTGCA 838
Qy 868 aaagtcaagggaagacgtgtcccgagcccaacgcagatcaagcgggcgccgccc 926
Db 839 AAGTCGGGGCAAGACGTGTCCAGGGCCCAACGCGCATCAAGCGGGGGGGGGCGCCCGC 897

RESULT 7
AL559806 938 bp mRNA EST 16-FEB-2001
LOCUS AL559806 LTI_FL011_BCl Homo sapiens cDNA clone CS0DG005YE05 5 prime
DEFINITION , mRNA sequence.
ACCESSION AL559806
VERSION EST.
KEYWORDS GI:12905654
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 938)
AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
```

```
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
Location/Qualifiers
1. 938
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DG005YE05"
/clone_lib="LTI_FL011_BCl"
/sex="male"
/tissue_type="B cells from Burkitt lymphoma"
/lab_host="DH10B"
/notes="Vector: pCMVSPORT 6; 1st strand cDNA was primed
with a NotI-oligo(dT) primer. Five prime end enriched,
double-stranded cDNA was digested with Not I and cloned
into the Not I and Eco RV sites of the pCMVSPORT 6 vector.
Library was constructed by Life Technologies. Contact :
Feng Liang Life Technologies, a division of Invitrogen
9800 Medical Center Drive Rockville, Maryland 20850, USA
Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com"
BASE COUNT 220 a 287 c 262 g 164 t 5 others
ORIGIN
Query Match 59.7% Score 824.6; DB 10; Length 938;
Best Local Similarity 96.6%; Pred. No. 1.9e-114;
Matches 906; Conservative 4; Mismatches 9; Indels 19; Gaps 6;
Qy 39 tgggggacaccttcacgtcacatcgccctgctgggctttgagaagcgtctcgaccga 98
Db 1 TGGGGGACACCTTCATCCGTACATCGCCCTGCTGGCTTTTGAGAGGCGCTTCGTACCCA 60
Qy 99 gcaagcactatgtatcatgttctctgtgaaatggcaggacctcgcgagaagtggtct 158
Db 61 GCCAGCACTA--TGTACATGTTCTGCTGTAATGGCAGGACCTGTCGAGAAGTGTCT 118
Qy 159 accgctctcaccgagatctacgagttccataaaaccttaaaagaagtgtccctattg 218
Db 119 ACCGGCGCTTCACCGAGATCTACGAGTTCCTAATAAACTTAAAGAAATGTTCCCTATTG 178
Qy 219 aggcaggggcgatcaatccagagacagaggatcatccccacacctccccagctcccaagtgt 278
Db 179 AGGCAGGGCGCATCAATCCAGAGAACAGGATCATCCCCACCTCCAGCTCCCAAGTGGT 238
Qy 279 ttgagggcgagggcgccgagaaacgcgagggcacactaccagtagtactgcagacgc 338
Db 239 TTGACGGGACGCGGGCGCGAGAACACCAGGACACTTACCGAGTACTGTCAGACGCG 298
Qy 339 tcattgacctgcccaccaaagatctcccgctgtccccacctctcgcagcttctcgaagtgc 398
Db 299 TCATGAGCCTGCCACCAAGATCTCCCGCTGTCCCGACCTCCTCGACTTCTTCAAGGTGC 358
Qy 399 gccctgatgacctcaagctcccccacggagacacaccagacacacacacacacacatacttga 458
Db 359 GCCCTGATGACCTCAAGCTCCCCACGGACACACAGAC--AAAAAGCCAGAGACATATTGA 417
Qy 459 tgcccaagatgcaagatcacgcgacacacatcacggcccccatcctcgcagacgt 518
Db 418 TGCCCAAGATGGCAAGATGATCCGCGACACATCACCGGCCCATCATCTCTRYAGACGT 477
Qy 519 accgcgccattgccactacgagaaagacctcgggctccgagatg-gctctgtccacgggg 577
Db 478 ACCGGCCATTGGCGACTACGAGAAAGCTCGGGCTCCGAGATGTGCTCTCTCCACGGGG 537
Qy 578 gacgtggtgaggtcgttagaagaagcgcgagcggttggtggttctcagatgaagacga 637
Db 538 GACGTGTGAGAGTCTGTGGAGAAAGAGCGAGACGCGNTTGGTGGTCTCTCAGATGAAGACA 597
Qy 638 aagcaggtggtatcccgagcgtctcctcctcgaccctcgacagctcctgcagacagcaa 697
Db 598 AAGCGAGCTGGATCCAGCATCTCTCTCGGCCCTTGACAGTCTCTGACGAGACGGAA 657
Qy 698 gacctgagcccaactatgcaggtgagccatacgtcgccatcaaggcctacactgctgtg 757
```

```

|||||
658 GACCCCTGAGCCCAACTATGAGGTGAGCCATACCTGCGCCATCAAGGCTACACTGCTGTG 717
Qy 758 gagggagagaggtgctccctgctcgaaggtgaagcttgaaggtcattcaaacctcgt 817
Db 718 GAGGGGACGAGGTGTCCCTGCTCGAGGGTGAAGCTGTGAGGTTCATTCACAAGCTCCTG 777
Qy 818 gacggct-----ggaagagagagctcacaggtcacttccctgctccatctacgt 865
Db 778 GACGGCTGGTGGTTCATCAGGAAGACGAGCTCACAGGCTACTTTCGTCATGTACCTG 837
Qy 866 caaagtcaagggaagcagctgtgtccaggtcccaacgagatcaagcgggggggcccgc 925
Db 838 CAAAGTCGGGGCAAGAYGTGTCCAGGCCCCAACGCCAGATCAAGCGGGGGGCGC--GCC 895
Qy 926 cgaaggtcgtccatccgcaacgagcagacatccacca 963
Db 896 CGCAGGTGCTCCATCCGCAAG-SCACAGCATCCATCA 932

RESULT 8
AL583422/c 895 bp mRNA EST 16-FEB-2001
LOCUS AL583422 LTI_NFL010_BC2 Homo sapiens cDNA clone CS0DL012YG22 3
DEFINITION prime, mRNA sequence.
ACCESSION AL583422
VERSION AL583422.1 GI:12952368
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 895)
AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
TITLE Full-length cDNA libraries and normalization.
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES
location/Qualifiers
source 1..895
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DL012YG22"
/clone_lib="LTI_NFL010_BC2"
/sex="male"
/tissue_type="B cells from Burkitt lymphoma"
/notes="vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the pCMVSPORT 6
vector. Library was normalized. Library was constructed by
Life Technologies. Contact : Feng Liang Life Technologies,
a division of Invitrogen 9800 Medical Center Drive
Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com"

BASE COUNT 131 a 272 c 292 g 198 t 2 others
ORIGIN

Query Match 58.9% Score 813.8; DB 10; Length 895;
Best Local Similarity 96.2%; Pred. No. 7.9e-113;
Matches 857; Conservative 2; Mismatches 19; Indels 13; Gaps 2;

Qy 436 aaaaagccagagacatacttgatgcccagaagtggaagtagtaccgagacagatcac 495
Db 894 AAAAGCCSAGAGACATACTTGATGCCCCAAGATGGCAAGAGTACCCGCGAGACATCAC 835
Qy 496 cggcccatcatctgcagacgtaccgcccattgccaaactacagagaacctgggtc 555
Db 834 CGGCCCATCATCTCTGCAGAGCTACCGCGCCATTGCGCGACTACGAGAAGACCTCGGGCTC 775

```

```

Qy 556 cgaatggctctgtccacgggggacgtggtaggttcgttagaagagcgagagcggtg 615
Db 774 CGAGATGGCTCTGTCCACGGGGGACGTGGTAGGTCTGGAGAGAGGAGACGGGTG 715
Qy 616 gtgattctgtatgataaagcaagcagagctggtatccacgcttcctctcgcagccct 675
Db 714 GTGTTTCTGTGATGAAAGCAAGCAGGAGGTGGATCCAGCATCTTCTCTCGAGCCCT 655
Qy 676 gacagtccttgacagagagaaagaccctgagcccaactatgcaggtgagccatcgtcgc 735
Db 654 GGACAGTCTCTGACGAGACGGGAAGACCTTGAGCCCACTATCGAGGTGAGCCATACGTGC 595
Qy 736 catcaagcctcacactgctgtgagggggagaggtgtccctgtctcaggttgaagctgt 795
Db 594 CATCAAGCCTTACACTGCTGTGGAGGGGACAGGTGTCCCTCTCGAGGGTGAAGCTGT 535
Qy 796 tgaggtcattcaagaagctctctggacggct-----ggaagagcagcgtcaagg 843
Db 534 TGAGGTCAATTCACAAGCTCTCTGGACGGCTGTGGGTATCATCAGGAAGACGACGTCA 475
Qy 844 ctactctccgtccatgtacctgcaaaagtcaggcgaagacgtgtccagggcccaagcca 903
Db 474 CTACTTTCCGTCATGTACTGTCAAAAGTCGGGCAAGACGTGTCCAGGCCCAACGCCA 415
Qy 904 gatcaagcgggggagcggcccccaggtgtccatccgcaacgcgcacagcatccacca 963
Db 414 GATCAAGCGGGGGCGCGCCCGCAGGTGTCTCCATCGCAACGCGCACAGCATCCATCA 355
Qy 964 gcgtctcgggaagcgtctcagcaggagcgtatctgcgcgaacagcgtctctttctgca 1023
Db 354 GCGGTCGCGGAAGGCGCTCAAGCAGGACGCGCTATCGCGCAACAGCGTCCGTTTTCTGCA 295
Qy 1024 gcagcagcggccagcagcggccggaacacagacagcccgagcggcgctcgagagga 1083
Db 294 GCAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 235
Qy 1084 gcgcagcagcagcgtctctaaacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1143
Db 234 GCGCAGCAGCGCAGCGCTCTAAGCGCAGCGCGGTGCCCGCGCGCGCGCGCGCGCG 175
Qy 1144 catctgaaccgtgcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1203
Db 174 CATCTTGAACCGTGCAGCGAGAGCACCACCAAGCGGAAGTGGCGTCTGCGCGTCTGAGG 115
Qy 1204 gacgcagtcctccagctagcgtctcgcccttgcgcgcgcgcgcgcgcgcgcgcgcgc 1263
Db 114 GAGCGCAGTCCCGAGTACGCTCTCGGCCCTTGCGGCCCTTGCGCGCGCGCGCGCGCG 55
Qy 1264 tatagcgtcgctgtgagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1313
Db 54 TATAGAGCCTAATCTTAACACCGAGRACACACCCGACCCCTATCCCAAC 4

RESULT 9
AL561079 906 bp mRNA EST 16-FEB-2001
LOCUS AL561079 LTI_NFL010_BC2 Homo sapiens cDNA clone CS0DL001YL02 5
DEFINITION prime, mRNA sequence.
ACCESSION AL561079
VERSION AL561079.1 GI:12908154
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 906)
AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
TITLE Full-length cDNA libraries and normalization.
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France

```

10	gcaaggagacactgagagccacccagtcattgggggacaccttaccgtcacatcgccct	69
11		
12		
13		
14		
15		
16		
17		
18		
19		
20		
21		
22		
23		
24		
25		
26		
27		
28		
29		
30		
31		
32		
33		
34		
35		
36		
37		
38		
39		
40		
41		
42		
43		
44		
45		
46		
47		
48		
49		
50		
51		
52		
53		
54		
55		
56		
57		
58		
59		
60		
61		
62		
63		
64		
65		
66		
67		
68		
69		
70	gctgggctttgagaagcgcttcgtaccacgcccagcactatgtgtacatgtttccctgggtaa	129
71		
72		
73		
74		
75		
76		
77		
78		
79		
80		
81		
82		
83		
84		
85		
86		
87		
88		
89		
90		
91		
92		
93		
94		
95		
96		
97		
98		
99		
100		
101		
102		
103		
104		
105		
106		
107		
108		
109		
110		
111		
112		
113		
114		
115		
116		
117		
118		
119		
120		
121		
122		
123		
124		
125		
126		
127		
128		
129		
130	atggcagaacctgcgagagaagtggtgtactaccggcgcttcacccagatctcacgaagtcca	189
131		
132		
133		
134		
135		
136		
137		
138		
139		
140		
141		
142		
143		
144		
145		
146		
147		
148		
149		
150		
151		
152		
153		
154		
155		
156		
157		
158		
159		
160		
161		
162		
163		
164		
165		
166		
167		
168		
169		
170		
171		
172		
173		

```
Db 62 CAGCCAGCACTATGTGTACATGTTCTCTGTAATGTCGAGGACCTGTTCGGAGAGGTGGT 121
Qy 157 ctaccggcgcttcaccgagatcagagttccataaaaccccttaaaagaatgttccctat 216
Db 122 CTACCGGGCGCTTCACCGGAGTCTACGAGTTCCATATAAACCTTAAAGAAATGTTCCCTAT 181
Qy 217 tgaggcaggggcgatcaatccagagaaacagagatcctcccccacccctccagctcccaagt 276
Db 182 TGAGCAGGGGGATCATCAGAGAACAGATCATCCCCACCTCCAGCTCCCAAGTG 241
Qy 277 gttgacggcgagcgggcgagaaacggcagggcacacttaccagtagtactgcagcac 336
Db 242 GTTTGACGGGACGCGGGCGCGAGAACCGCAGGACACTTACCGAGTAGTCTGCAGCAC 301
Qy 337 gctcatgagctgcggcacaagatctccgctgttcccaactcctcgtactcttcaaggt 396
Db 302 GCTCATGAGCTTCCGCCACCAAGATCTCCGCTGTCCCACTCTCCGACTTCTTCAAGGT 361
Qy 397 ggcctctgatgacctcaagctcccccacggacaaacagacacaaagccagagacatactt 456
Db 362 GGGCCTGATGACCTCAAGCTCCCCACGACAAACAGACACAAAAAGCCAGACATACTT 421
Qy 457 gatgcccagaagatgcaagagtagtaccgacagacatcccccacccatcctctgcagac 516
Db 422 GATGCCAAAGATGCAAGATACCGGACAGACATCACCGGCCCATCATCCTCTGCAGAC 481
Qy 517 gtaccgcccattgcgaactcagagaagacctcggtctccagatgctctgcccagg 576
Db 482 GTACCGGCCCATTTGCCACTACGAGAAACCTTCGGGCTCCGAGATGGCTCTGTCCACGGG 541
Qy 577 ggaactggtgaggt-cgtagagaagagcagagcgggtggtggtctgtcagatgaag 635
Db 542 GCACCTGTGGAGTCCGTAGAGAGAGCGAGAGCGGTTGTGTCTGTCTGATGAAG 601
Qy 636 caaagcagggtggtatccagcaggtctctctcagagccctggaagctcctcaaga-gacg 694
Db 602 CAAAGCAGGCTGGATCCAGCGCTTTCTCCAGGCCCTGGACAGTCTCTCAGCAGGACG 661
Qy 695 gaagacctgagcccaact--atgcagtgagccatagc-tcgccatcaagggctacact 751
Db 662 GAAGACCTTGAGCCCACTTATGAGGGTGAGCCATAGTTTCGCCATCAAGGCTTACACT 721
Qy 752 gctgtggggggagaggttccctcgtcagaggtgaagctgttgaggtcattcacaag 811
Db 722 GCTGTGTAGGGGAACGAGGTGTCCTCTGTCGAGGTGAAGCTGTGAAGTCAATTTCACAAG 781
Qy 812 ctctggagcg 821
Db 782 CTCTGGGACG 791

RESULT 15
AL581347/c
LOCUS AL581347 LTI_FL011_BCl Homo sapiens cDNA clone CS0DG005YE05 3 prime
DEFINITION , mRNA sequence.
ACCESSION AL581347
VERSION AL581347.1 GI:12948254
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 834)
AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES
Location/Qualifiers
1..834
source
```

```
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="LTI_FL011_BCl"
/sex="male"
/tissue_type="B cells from Burkitt lymphoma"
/lab_host="DH10B"
/note="Vector: pCMVSPORT 6; 1st strand cDNA was primed
with a NotI-oligo(dT) primer. Five prime end enriched,
double-stranded cDNA was digested with Not I and cloned
into the Not I and Eco RV sites of the pCMVSPORT 6 vector.
Library was constructed by Life Technologies. Contact :
Feig Liang Life Technologies, a division of Invitrogen.
9800 Medical Center Drive Rockville, Maryland 20850, USA
Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com"
BASE COUNT 126 a 238 c 271 g 191 t 8 others
ORIGIN
Query Match 51.5%; Score 712.2; DB 10; Length 834;
Best Local Similarity 95.0%; Pred. No. 1.3e-97;
Matches 777; Conservative 5; Mismatches 19; Indels 17; Gaps 4;
Qy 318 ttaccgagtagtcagcacgctcatgagcctgcccacacaaatctccgctgtcccaacc 377
Db 816 TTACCGAGTACTGCAS-ACGCTCATGACCTGCCACCAAGATCTCCGCTGTCCCAACC 758
Qy 378 tctcgtactcttcaagtgcgccctgatgacctcaagctcccccacggacacacagacaa 437
Db 757 TCCTCGACTTCTTCAAGGTGGCGCTGATGACCTCAAGTCTCCACGACACACAGAC-A 699
Qy 438 aaaagcagaagatacttgatgccaaaagatggcagagtagtaccgacagacataccag 497
Db 698 AAAAGCCAGAGACATACTTGTATGCCAAAGATGGCAAGTACCGCGACAGACATCACC 639
Qy 498 gcccatcatcctgcagacgtaccgcccattgcacactacgagaagacctcgagctccg 557
Db 638 GCCCATCATCTCTCGACAGCTACCGCGCCATTTGCCGACTACGAGAAGACCTTCGGGCTCCG 579
Qy 558 agatggctctgtccacgggggacgtggtggaggtcgtagaagagcagagcggttgggt 617
Db 578 AGATGGCTCTGTCCACGGGGACGCTGGTGGAGGTCTGTGGAGAGAGAGAGCGGTTGGT 519
Qy 618 ggttctcgaataaaagcagaagcgtgatcccaagctccttctcgtgagccctgg 677
Db 518 GGTTCGTGTCAGATGAAGCAAGCAGGCTGATTCCTCCAGCATCTCTCTCGAGGCCCTGG 459
Qy 678 acagtctgcagagacgaaagacctgagcccaactatgcaggtgagccatacagtcgcca 737
Db 458 ACAGTCTGACGAGACGGAAGACCTTGAGCCCACTATGCAAGGTGAGCCATACGTCGCCA 399
Qy 738 tcaaggctacactgctgtggagggggacgaggtgtccctgctcgtgaggttgaagctgtg 797
Db 398 TCAAGGCTTACACTGCTGTGGAGGGGACGAGGTGTCCCTGCTCGAGGKTGAAGCTGTG 339
Qy 798 aggtcattcacaagctcctgagcgcctgg-----aaagacgagcgtcacagggc 844
Db 338 AGKTCATTCACAAGCTCTCTGACCGCTGGTGGGTTCATCAGGAAAGAAACATACACAGGC 279
Qy 845 tacttccgctcatgtacctgcaaaagtccagggcgaagcgtgtcccgagggcccaacgcag 904
Db 278 TACTTTCCGTCCATGTACCTTGCAAAAGTCGGGGCAAGACGCTGTCCAGGGCCCAACGCCAG 219
Qy 905 atcaagcggggg--gcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 962
Db 218 ATCAAGCGGTGTGKTGCGCGCGCGCGCGCAGGTGCTCATTCGCAACGCGCAACATCATC 159
Qy 963 agcgttcgcggaagcgcctcagccagcagccctatccgcgaacagcgtccgttttctgc 1022
Db 158 AGCGGTGCGGAAGCGCTCAGCCAGGACSCCTATCCTCCGCAACAGCGTCTCGTTTCTGC 99
Qy 1023 agcagcagcgcgcgcagggcgcgccgggacgcagagccccgggagccccgctcagaggg 1082
```


THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 11, 2002, 09:07:28 ; Search time 814.02 Seconds
(without alignments)
1455.521 Million cell updates/sec

Title: US-09-820-005-1
Perfect score: 1382
Sequence: 1 cctggaagtgcaggagca.....aaaaaaaaaaaaaaaa 1382

Scoring table: IDENTITY_NUC
Gap 10.0, Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_1101:
1: /SID88/gcgdata/geneseq/geneseq/NA1980.DAT.*
2: /SID88/gcgdata/geneseq/geneseq/NA1981.DAT.*
3: /SID88/gcgdata/geneseq/geneseq/NA1982.DAT.*
4: /SID88/gcgdata/geneseq/geneseq/NA1983.DAT.*
5: /SID88/gcgdata/geneseq/geneseq/NA1984.DAT.*
6: /SID88/gcgdata/geneseq/geneseq/NA1985.DAT.*
7: /SID88/gcgdata/geneseq/geneseq/NA1986.DAT.*
8: /SID88/gcgdata/geneseq/geneseq/NA1987.DAT.*
9: /SID88/gcgdata/geneseq/geneseq/NA1988.DAT.*
10: /SID88/gcgdata/geneseq/geneseq/NA1989.DAT.*
11: /SID88/gcgdata/geneseq/geneseq/NA1990.DAT.*
12: /SID88/gcgdata/geneseq/geneseq/NA1991.DAT.*
13: /SID88/gcgdata/geneseq/geneseq/NA1992.DAT.*
14: /SID88/gcgdata/geneseq/geneseq/NA1993.DAT.*
15: /SID88/gcgdata/geneseq/geneseq/NA1994.DAT.*
16: /SID88/gcgdata/geneseq/geneseq/NA1995.DAT.*
17: /SID88/gcgdata/geneseq/geneseq/NA1996.DAT.*
18: /SID88/gcgdata/geneseq/geneseq/NA1997.DAT.*
19: /SID88/gcgdata/geneseq/geneseq/NA1998.DAT.*
20: /SID88/gcgdata/geneseq/geneseq/NA1999.DAT.*
21: /SID88/gcgdata/geneseq/geneseq/NA2000.DAT.*
22: /SID88/gcgdata/geneseq/geneseq/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	384	27.8	384	22	AAH45277
2	323.8	23.4	425	21	AAC98438
3	154.8	11.2	192	22	AAH45280
4	99.8	7.2	3195	22	AAS22715
5	88	6.4	5588	22	AAS22479
6	77.2	5.6	7368	21	AAF21398
7	77.2	5.6	11651	21	AAF21397
8	75	5.4	75	21	AAF21397
9	75	5.4	75	21	AAF21397
10	52.4	3.8	3201	21	AAH45275
11	51.6	3.7	1214	21	AAC99095

c	12	49	3.5	1305	22	AAH45277
c	13	49	3.5	1305	22	AAH73513
c	14	49	3.5	1316	21	AAH73513
c	15	48.6	3.5	114955	20	AAH53491
	16	47.8	3.5	1803	20	AAH89760
	17	47.8	3.5	1850	21	AAH46151
	18	47.8	3.5	1858	20	AAH89759
	19	47.8	3.5	1974	21	AAH71172
	20	46.4	3.4	1635	21	AAH59474
	21	46.4	3.4	2115	20	AAH32263
	22	46.4	3.4	2115	21	AAH88198
	23	46.4	3.4	2205	22	AAH08194
	24	46	3.3	734	19	AAH59738
	25	46	3.3	1513	11	AAH06160
	26	45.8	3.3	570	21	AAH65103
	27	45.8	3.3	570	22	AAH44249
	28	45.8	3.3	1163	14	AAH35198
	29	45.8	3.3	1541	14	AAH35198
	30	45.6	3.3	1043	21	AAH02075
	31	45.6	3.3	1050	21	AAH298064
	32	45.6	3.3	1050	22	AAH11677
	33	45	3.3	1233	21	AAH21869
	34	44.6	3.2	10732	21	AAH10594
	35	44.4	3.2	2004	17	AAH13952
	36	44.4	3.2	2004	18	AAH73997
	37	44.4	3.2	2004	19	AAH16170
	38	44.4	3.2	2010	15	AAH74683
	39	44.4	3.2	2576	17	AAH13945
	40	44.4	3.2	2576	18	AAH74006
	41	44.4	3.2	2576	19	AAH16181
	42	44.4	3.2	2655	15	AAH74682
	43	44.4	3.2	2655	17	AAH13951
	44	44.4	3.2	2655	17	AAH13954
	45	44.4	3.2	2655	18	AAH73996

ALIGNMENTS

RESULT 1
AAH45277
ID AAH45277 standard; DNA; 384 BP.
XX
AC AAH45277:
XX
DT 07-SEP-2001 (first entry)
XX
DE DNA encoding human PX domain.
XX
Human; antiinflammatory; protein coordinate data; chemical shift data;
KW nuclear magnetic resonance; NMR; structural coordinate data; PX domain;
KW reduced nicotinamide adenine dinucleotide phosphate; NADPH;
KW NADPH oxidase; superoxide production; inflammatory disease; ds.
XX
OS Homo sapiens.
XX
PN WO200142453-A1.
XX
PD 14-JUN-2001.
XX
PF 01-DEC-2000; 2000WO-JP08501.
XX
PR 06-DEC-1999; 99JP-0346193.
XX
PA (BIOM-) BIOMOLECULAR ENG RES INST.
XX
PI Kohda D, Hiroaki H, Sumimoto H;
XX
DR WPI; 2001-381679/40.
XX
PT Regulation of function of a protein containing a PX domain for
PT controlling an inflammatory response to disease or trauma, comprises
PT using nuclear magnetic resonance (NMR) chemical shift and structural


```
Db 343 aaagatgcaagatcgcgcagacatcacccgccccatcctcgcagagtcaccgc 402
QY 524 gccattgccactacgagaagac 546
Db 403 gccattgccactacgagaagac 425

RESULT 3
AAH45280
ID AAH45280 standard; DNA: 192 BP.
XX AC AAH45280;
XX DT 07-SEP-2001 (first entry)
XX DE DNA encoding human SH3 domain.
XX KW Human; antiinflammatory; protein coordinate data; chemical shift data;
KW nuclear magnetic resonance; NMR; structural coordinate data; PX domain;
KW reduced nicotinamide adenine dinucleotide phosphate; NADPH; SH3 domain;
KW NADPH oxidase; superoxide production; inflammatory disease; ds.
XX OS Homo sapiens.
XX PN WO200142453-A1.
XX PD 14-JUN-2001.
XX PF 01-DEC-2000; 2000WO-JP08501.
XX PR 06-DEC-1999; 99JP-0346193.
XX PA (BIOM-) BIOMOLECULAR ENG RES INST.
XX PI Kohda D, Hiroaki H, Sumimoto H;
XX WPI; 2001-381679/40.
XX PT Regulation of function of a protein containing a PX domain for
XX controlling an inflammatory response to disease or trauma, comprises
XX using nuclear magnetic resonance (NMR) chemical shift and structural
XX coordinate data.
XX PS Example 2; Page 186; 195pp; Japanese.
XX CC The present sequence is provided in a specification relating to the use
XX of chemical shift data from nuclear magnetic resonance (NMR) and
XX structural coordinate data of the PX domain of a protein to search for,
XX evaluate, design and identify variant PX domain sequences for
XX controlling the function of proteins containing the PX domain. It also
XX relates to compounds promoting the binding of substances to the PX
XX domain, and compounds inhibiting the binding of substances to the PX
XX domain. The invention can be used in the identification of substances
XX that can be used in the regulation of reduced nicotinamide adenine
XX dinucleotide phosphate (NADPH) oxidase activity, which is involved in
XX superoxide production in inflammatory response in disease or trauma.
XX SQ Sequence 192 BP; 44 A; 50 C; 61 G; 37 T; 0 other;

Query Match 11.2%; Score 154.8; DB 22; Length 192;
Best Local Similarity 92.7%; Pred. No. 6.8e-24;
Matches 178; Conservative 0; Mismatches 2; Indels 12; Gaps 1;

QY 704 gagcccaactatgcagtgagccatcgcgcacacagccctacactcgtgtggaagg 763
Db 1 gagcccaactatgcagtgagccatcgcgcacacagccctacactcgtgtggaagg 60
QY 764 gagaggtgtccctgctcgcaggggtgaagctgttgaggtcattcacagctcctggagcgc 823
Db 61 gagaggtgtccctgctcgcaggggtgaagctgttgaggtcattcacagctcctggagcgc 120
```

```
QY 824 t-----ggaaagacgacgtcacagagctactccctccatctccctgcaaaag 871
Db 121 tgggtggtcatcagaaaagacgacgtcacaggtacttccgtccatgtacctgcaaaag 180
QY 872 tcagggcgaagac 883
Db 181 tcggggcgaagac 192

RESULT 4
AAS22715
ID AAS22715 standard; cDNA: 3195 BP.
XX AC AAS22715;
XX DT 24-OCT-2001 (first entry)
XX DE Human cDNA encoding a novel human protein #281.
XX KW Human; novel protein; ss; Antianaemic; osteopathic; antiinflammatory;
KW immunomodulatory; cytostatic; neuroprotective; vulnery; nootropic;
KW anticonvulsant; antiarthritic; cerebroprotective; antifungal; antiviral;
KW antibacterial; antiallergic; dermatological; haemostatic; antiasthmatic;
KW thrombolytic; immunogen; antibody; gene therapy; neurological disorder;
KW Parkinson's disease; inflammatory disorder; cancer; asthma; osteoporosis;
KW tissue regeneration; immune disorder.
XX OS Homo sapiens.
XX PN WO200155437-A2.
XX PD 02-AUG-2001.
XX PF 25-JAN-2001; 2001WO-US02623.
XX PR 25-JAN-2000; 2000US-0491404.
XX PA (HYSE-) HYSEQ INC.
XX PI Tang YT, Liu C, Drmanac RT;
XX WPI; 2001-451939/48.
XX DR P-PSDB; AAU14410.
XX PT Isolated polypeptides useful for treating anti-inflammatory diseases,
XX nervous system disorders, and for regenerating bone and cartilage.
XX CC Claim 1; Page 694-695; 894pp; English.
XX CC The invention relates to polynucleotides encoding novel human
XX proteins or their active domains. The polypeptides, polynucleotides and
XX antibodies raised against the polypeptides are used in a method of
XX treatment of a mammal and prevention of disorders caused by the aberrant
XX protein expression or activity. The polypeptides can be used as
XX molecular weight markers, food supplements, and in antibody production.
XX The polypeptides are used to identify compounds which bind to the
XX polypeptides. Polynucleotides of the invention are used as probes and
XX primers, for sequencing, for chromosome or gene mapping, in the
XX production of recombinant proteins, and in generating anti-sense DNA or
XX RNA and in gene therapy. Polypeptides of the invention can be used to
XX target drugs to a tumour, in assays to determine biological activity, to
XX raise antibodies/elicite an immune response, to determine quantitative
XX protein levels, as tissue markers, and to isolate receptors or ligands.
XX Polypeptides of the invention may also be useful in treating platelet
XX disorders, stem cell disorders, regenerating bone, cartilage, tendon,
XX ligament and/or nerve tissue, wound healing, treating burns, promoting
XX the proliferation, differentiation and survival of stem cells, as a
XX contraceptive, treating osteoporosis and osteoarthritis, anaemia,
XX Alzheimer's, Parkinson's and Huntington's diseases, amyotrophic lateral
XX sclerosis, stroke, immune deficiencies resulting from bacterial, viral or
XX fungal infection or from autoimmunity, cancer, allergy, asthma,
XX graft-versus-host disease, eczema, haemophilia, thrombosis,
XX anti-inflammatory diseases, nervous system disorders, and infection.
```


KW respiratory obstruction; pulmonary obstruction; impeded respiration;
 KW surfactant hypoproduction; pulmonary vasoconstriction; asthma; RDS;
 KW respiratory distress syndrome; pain; cystic fibrosis; allergic rhinitis;
 KW pulmonary hypertension; emphysema; pulmonary transplantation rejection;
 KW chronic obstructive pulmonary disease; pulmonary infection; bronchitis;
 KW cancer; ss.

OS Homo sapiens.

PN WO200062736-A2.

PD 26-OCT-2000.

24-MAR-2000; 2000WO-US08020.

PR 06-APR-1999: 99US-0127958-

PA (UYEC-) UNIV EAST CAROLINA.

PA (NYCE/) NYCE J W.

PI Nyce JW;

DR WPI; 2000-679539/66.

Low adenosine (A) content antisense oligonucleotides which do not trigger adenosine receptors during metabolism, useful e.g. for treating cancers and respiratory obstructions -

PS Disclosure; Page 1383-1385; 1592pp; English.

The present invention describes low adenosine (A) content antisense oligonucleotides and compositions (I) comprising them. In the antisense oligonucleotides the A is replaced by a 'Universal' or alternative base. (I) can have respiratory, bronchodilator, antiinflammatory, analgesic, immunosuppressive, antiasthmatic, hypotensive and cytostatic activities. The antisense oligonucleotides and (I) can be used to down-regulate the expression and/or activity of target polypeptides associated with lung/respiratory disorders and malignancies, such as stimulating and activating peptide factors and transmitters, transcription factors, immunoglobulins and antibodies, antibody receptors, cytokines and chemokines, endogenously produced specific and non-specific enzymes, binding proteins, adhesion molecules and their receptors, cytokine and chemokine receptors, adenosine receptors, bradykinin receptors, central nervous system (CNS) and peripheral nervous and non-nervous system receptors, CNS and peripheral nervous and non-nervous system peptide transmitters, defensins, growth factors, vasoactive peptides and receptors, binding proteins and malignancy associated proteins. The antisense oligonucleotides may be used in this way to treat disorders including respiratory obstruction (especially pulmonary obstruction and/or bronchoconstriction) and/or lung inflammation, allergy(ies) and/or surfactant hypoproduction which are associated with a disease or condition selected from pulmonary vasoconstriction, inflammation, allergies, asthma, impeded respiration, respiratory distress syndrome (RDS), pain, cystic fibrosis (CF), allergic rhinitis (AR), pulmonary hypertension, emphysema, chronic obstructive pulmonary disease (COPD), pulmonary transplantation rejection, pulmonary infections, bronchitis, and/or cancer. AAR18434 to AAR21543 represent human polynucleotide fragments and antisense oligonucleotides used in the exemplification of the present invention.

Sequence 7368 BP: 2018 A: 1656 C: 1890 G: 1804 T: 0 other:

Query Match	5.6%	Score 77.2;	DB 21;	Length 7368;
Best Local Similarity	96.3%	Pred. No. 3.4e-07;		
Matches	79;	Conservative	0;	Mismatches 3;
				Indels 0;
				Gaps 0;

50 ttcatccgtcacatcgccctgctgagctttgagaagcgccttcgtacccagccagcactat 109

db 7287 ttccgtaaagtccacatcggccctgctgggctttgagaaaggccttcgtaccagccagcactat 7346

Qy 110 gtgtacatgttcctcgtggaat 131

Db 7347 atgtacatatattcctgaatgaaat 7368

RESULT

AAA35276

ID AAA35276 standard; DNA; 11651 BP.

AC AAA35276;

DT 28-JUL-2000 (first entry)

DE Human adenosine receptor related polynucleotide 2nd SEQ ID NO:150.

Human: adenosine receptor; low adenosine infusate oligonucleotide; phosphorothioate; impaired respiration; inflammation; allergy; allergic disease; bronchoconstriction; inhibitor; antiinflammatory; antiallergic; antiasthmatic; cytostatic; analgesic; impaired airway; lung disease; ischaemic condition; pulmonary vasoconstriction; asthma; respiratory distress syndrome; pain; cystic fibrosis; emphysema; pulmonary hypertension; chronic obstructive pulmonary disease; COPD; cancer; leukaemia; lymphoma; carcinoma; metastasis; ss.

OS Homo sapiens.

AA PN W0200009525-A2

24-FEB-2000

AA
PF 03-AUG-1999; 99WO-US17712.

03-AUG-1998 090521Z

XX PA (UUEC-) IINTV EAST CAROLINA

XX
PI Nvce JW:

WPI: 2000-205971/18.

PT New antisense oligonucleotides useful for treating e.g. pulmonary
PT vasoconstriction, inflammation, allergies, asthma, hypertension,
PT bronchitis, emphysema, respiratory distress syndrome, ischemia or
PT cancer

XX
PS Disclosure: Page 1298-1301: 1343nn: English

The present invention describes a new composition comprising an antisense oligonucleotide (ON) with low adenosine (up to 15%), which targets nucleic acids involved in bronchoconstriction, allergies, and/or inflammation. The ON can have antiinflammatory, antiallergic, antisthmatic, cytostatic and analgesic activities. The compositions are useful for the treatment of diseases associated with inflammation, impaired airways, including lung disease and diseases whose secondary effects afflict the lungs of a subject. They can be used for treating e.g. ischaemic conditions, pulmonary vasoconstriction, allergies, asthma, impeded respiration, respiratory distress syndrome, pain, cystic fibrosis, pulmonary hypertension, emphysema, chronic obstructive pulmonary disease (COPD), and cancers such as leukaemias, lymphomas, carcinomas, and cancers which may metastasize to the lungs, including breast and prostate cancer. The reduction of the adenosine content of the ONs reduces side effects. The A-containing ONs break down with the release of deoxyadenosine which activates adenosine receptors causing bronchoconstriction and inflammation. AAA32313 to AAA35312 represent nucleotide sequences given in the sequence listing from the present invention, which correspond to SEQ ID NO:1 to 2815, and then the last 195 sequences are also called SEQ ID NO:1 to 185, but the sequences differ from the previously named sequences. SEQ ID NO:11 to 1680 (AAA32323 to AAA33992) are specifically claimed ONs from the present invention. N.B. Sequences given in the disclosure of the present invention do not match up with their corresponding SEQ ID NO: sequences given in the sequence listing.

Sequence 11651 BP: 3079 A: 2635 C: 2989 G: 2948 T: 0 other: XX

Query Match 5.6%; Score 77.2; DB 21; Length 11651;
Best Local Similarity 96.3%; Pred. No. 3.7e-07;
Matches 79; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 50 ttcatcgtcacatcgccctgctgggttttgagaagcgcttcgtaccacgcagcaactat 109
||| | ||||||||||||||||||||||||||||||||||||||||||||||||||
Db 11570 ttgatgtcacatcgccctgctgggttttgagaagcgcttcgtaccacgcagcaactat 11629

QY 110 gtgtacatgttctctggtgaat 131
||||||||||||||||||||||||||||||||||||||||||||||||
Db 11630 gtgtacatgttctctggtgaat 11651

RESULT 8
AAF21397
ID AAF21397 standard; DNA; 75 BP.
XX
AC AAF21397;
XX
DT 14-MAR-2001 (first entry)
XX
DE Human low adenosine antisense oligonucleotide related sequence #2964.
XX
KW Low adenosine antisense oligonucleotide; phosphorothioate; allergy;
human; airway disorder; bronchoconstriction; lung inflammation;
surfactant depletion; respiratory; bronchodilator; antiinflammatory;
immunosuppressive; antiasthmatic; analgesic; hypotensive; cytostatic;
respiratory obstruction; pulmonary obstruction; impeded respiration;
surfactant hypoproduction; pulmonary vasoconstriction; asthma; RDS;
respiratory distress syndrome; pain; cystic fibrosis; allergic rhinitis;
pulmonary hypertension; emphysema; pulmonary transplantation rejection;
chronic obstructive pulmonary disease; pulmonary infection; bronchitis;
cancer; ss.
XX
OS Homo sapiens.
XX
PN WO200062736-A2.
XX
PD 26-OCT-2000.
XX
XX 24-MAR-2000; 2000WO-US08020.
XX
PF 06-APR-1999; 99US-0127958.
XX
PR {UYEC-} UNIV EAST CAROLINA.
(NYCE/) NYCE J W.
XX
PI Nyce JW;
XX
PP WPI; 2000-679539/68.
XX
PT Low adenosine (A) content antisense oligonucleotides which do not
trigger adenosine receptors during metabolism, useful e.g. for treating
cancers and respiratory obstructions -
XX
PS Disclosure; Page 1383; 1592pp; English.
XX

CC The present invention describes low adenosine (A) content antisense
CC oligonucleotides and compositions (I) comprising them. In the antisense
CC oligonucleotides the A is replaced by a 'Universal' or alternative base.
CC (I) can have respiratory, bronchodilator, antiinflammatory, analgesic,
CC immunosuppressive, antiasthmatic, hypotensive and cytostatic activities.
CC The antisense oligonucleotides and (I) can be used to down-regulate the
CC expression and/or activity of target polypeptides associated with
CC lung/respiratory disorders and malignancies, such as stimulating and
CC activating peptide factors and transmitters, transcription factors,
CC immunoglobulins and antibodies, antibody receptors, cytokines and
CC chemokines, endogenously produced specific and non-specific enzymes,
CC binding proteins, adhesion molecules and their receptors, cytokine and
CC chemokine receptors, adenosine receptors, bradykinin receptors, central
CC nervous system (CNS) and peripheral nervous and non-nervous system
CC • receptors, CNS and peripheral nervous and non-nervous system peptide

transmitters, defensins, growth factors, vasoactive peptides and receptors, binding proteins and malignancy associated proteins. The antisense oligonucleotides may be used in this way to treat disorders including respiratory obstruction (especially pulmonary obstruction CC and/or bronchoconstriction) and/or lung inflammation, allergy(ies) CC and/or surfactant hypoproduction which are associated with a disease or CC condition selected from pulmonary vasoconstriction, inflammation, CC allergies, asthma, impeded respiration, respiratory distress syndrome (RDS), pain, cystic fibrosis (CF), allergic rhinitis (AR), pulmonary CC hypertension, emphysema, chronic obstructive pulmonary disease (COPD), CC pulmonary transplantation rejection, pulmonary infections, bronchitis, CC and/or cancer. AAF18434 to AAF21543 represent human polynucleotide CC fragments and antisense oligonucleotides used in the exemplification of CC the present invention.

XX
SQ Sequence 75 BP; 15 A; 21 C; 19 G; 20 T; 0 other;

Query Match 5.4%; Score 75; DB 21; Length 75;
Best Local Similarity 100.0%; Pred. No. 3.7e-07;
Matches 75; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 57 gtacatcgccctgctgggtttgagaagcgcttcgaccagcgacactatgtgtaca 116
Db 1 gtacatcgccctgctgggtttgagaagcgcttcgaccagcgacactatgtgtaca 60
QY 117 tgttcctgggtgaaat 131
Db 61 tgttcctgggtgaaat 75

RESULT 9
ID AAA35275
AA ID AAA35275 standard; DNA; 75 BP.
XX
AC AAA35275;
XX
DT
XX 28-JUL-2000 (first entry)
XX
DE Human adenosine receptor related polynucleotide 2nd SEQ ID NO:149.
XX
KW Human; adenosine receptor; low adenosine antisense oligonucleotide;
KW phosphothioate; impaired respiration; inflammation; allergy;
KW allergic disease; bronchoconstriction; inhibitor; antiinflammatory;
KW antiallergic; antilasthmatic; cytostatic; analgesic; impaired airway;
KW lung disease; ischaemic condition; pulmonary vasoconstriction; asthma;
KW respiratory distress syndrome; pain; cystic fibrosis; emphysema;
KW pulmonary hypertension; chronic obstructive pulmonary disease; COPD;
KW cancer; leukaemia; lymphoma; carcinoma; metastasis; ss.
XX
OS Homo sapiens.
XX
PN WO200009525-A2.
XX
PD 24-FEB-2000.
XX
PF 03-AUG-1999; 99WO-USI7712.
XX
PR 03-AUG-1998; 98US-0095212.
XX
PA (UYEC-) UNIV EAST CAROLINA.
XX
PI Nyce JW;
XX
DR WPI; 2000-205971/18.
XX
PT New antisense oligonucleotides useful for treating e.g. pulmonary
PT vasoconstriction, inflammation, allergies, asthma, hypertension,
PT bronchitis, emphysema, respiratory distress syndrome, ischemia or
PT cancers
XX
PS Disclosure; Page 1298; 1343pp; English.
XX

AAC99095
ID AAC99095 standard; cDNA: 1214 BP.
AC AAC99095;
XX
XX
DT 09-MAR-2001 (first entry)
DE Human pancreatic cancer antigen nucleotide sequence SEQ ID NO:323.
XX
XX
KW Human: pancreas; pancreatic cancer; pancreatic cancer antigen;
KW detection; diagnosis; identification; cytostatic; neuroprotective;
KW nontropic; immunomodulatory; relaxant; contraceptive; gynaecological;
KW antiinflammatory; cardiant; gene therapy; chromosome mapping;
KW linkage analysis; tissue identification; tissue typing; forensic;
KW neural; immune system; muscular; reproductive; gastrointestinal;
KW pulmonary; cardiovascular; renal; proliferative; ss.
XX
OS Homo sapiens.
XX
PN WO200055320-A1.
XX
PD 21-SEP-2000.
XX
PF 08-MAR-2000; 2000WO-US05989.
XX
PR 12-MAR-1999; 99US-0124270.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Ruben SM;
XX
DR WPI: 2000-579444/54.
DR P-PSDB; AAB54330.
XX
PT New nucleic acid that is a pancreatic cancer antigen for preventing,
PT treating, or ameliorating a medical condition, particular pancreatic
PT cancer, or for use in assays for diagnosing a pathological condition -
XX
PS Claim 1; Page 748; 1379pp; English.
XX
CC AAC98773 to AAC99231 encode the human pancreatic cancer associated
CC proteins, called pancreatic cancer antigens, given in AAB54008 to
CC AAB54466. The human pancreatic cancer antigens have cytostatic,
CC neuroprotective, nontropic, immunomodulatory, relaxant, contraceptive,
CC gynaecological, cardiant and antiinflammatory activities, and can be used
CC in gene therapy. The polynucleotide and proteins can be used for
CC preventing, treating, or ameliorating a medical condition or in assays
CC for diagnosing a pathological condition or a susceptibility to one in a
CC subject. Binding partners to the proteins and the activity of the
CC proteins can be identified. The pancreatic cancer antigens can be used to
CC detect, treat or prevent pancreatic disorders, especially cancer.
CC Agonists and antagonists to the antigens can be screened for. The
CC pancreatic cancer antigen polynucleotides can be used to design nucleic
CC acid hybridisation probes that can be used in chromosome mapping, linkage
CC analysis, tissue identification and/or typing and a variety of forensic
CC and diagnostic methods. The proteins can be used to generate antibodies
CC which are used to purify, detect and target the polypeptides, including
CC both in vivo and in vitro diagnostic and therapeutic methods. The
CC proteins can be used to treat or prevent neural, immune system, muscular,
CC reproductive, gastrointestinal, pulmonary, cardiovascular, renal or
CC proliferative disorders. AAC99232 to AAC99240 and AAB54467 represent
CC sequences used in the exemplification of the present invention.
XX
SQ Sequence 1214 BP; 264 A; 395 C; 325 G; 229 T; 1 other;

Query Match 3.7%; Score 51.6; DB 21; Length 1214;
Best Local Similarity 58.4%; Pred. No. 0.057;
Matches 90; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

Qy 1229 ggccttcgcccgcgcctgtatctatcgctgtctctatagacgtgcgtctgacgcgcg 1288
Db 1040 gtcacttcgctgcctgcgtctctgtgtgtgtgagcagaagtggagctggggggcg 1099

Qy 1289 agggcagcccccagccctgtccagcgcgtccgcacccccaataatgttcttgga 1348
Db 1100 tggagagcccgccgcctgcacccctccctgacccgtctatataataatagagtgt 1159
Qy 1349 gtggaataaaaaaataaaaaaataaaaaa 1382
Db 1160 gtctaaaaaataaaaaaataaaaaa 1193

RESULT 12
AAS07945/c
ID AAS07945 standard; cDNA: 1305 BP.
XX
AC AAS07945;
XX
DT 23-OCT-2001 (first entry)
XX
DE Human cDNA encoding G-protein coupled receptor, hRUP18.
XX
KW Human; G-protein coupled receptor; GPCR; hRUP18; agonist;
KW inverse agonist; lung cancer; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 1..1305
FT FT /*tag= a
FT FT /product= "hRUP18"
XX
PN WO200136471-A2.
XX
PD 25-MAY-2001.

XX 16-NOV-2000; 2000WO-US31509.
XX 17-NOV-1999; 99US-0166088.
XX 17-NOV-1999; 99US-0166099.
XX 17-NOV-1999; 99US-0166369.
XX 23-DEC-1999; 99US-0171900.
XX 23-DEC-1999; 99US-0171901.
XX 23-DEC-1999; 99US-0171902.
XX 11-FEB-2000; 2000US-0181749.
XX 14-MAR-2000; 2000US-0189258.
XX 14-MAR-2000; 2000US-0189259.
XX 10-APR-2000; 2000US-0195898.
XX 10-APR-2000; 2000US-0195899.
XX 28-APR-2000; 2000US-0196078.
XX 12-MAY-2000; 2000US-0200419.
XX 12-MAY-2000; 2000US-0203630.
XX 12-JUN-2000; 2000US-0210741.
XX 12-JUN-2000; 2000US-0210982.
XX 21-AUG-2000; 2000US-0226760.
XX 26-SEP-2000; 2000US-0235418.
XX 26-SEP-2000; 2000US-0235779.
XX 20-OCT-2000; 2000US-0242332.
XX 20-OCT-2000; 2000US-0242343.
XX (AREN-) ARENA PHARM INC.
XX Chen R, Dang HT, Lowitz KP;
PI P-PSDB; AAU04372.
XX WPI: 2001-355616/37.
DR P-PSDB; AAU04372.
XX
XX Endogenous and non-endogenous versions of human G-protein coupled
XX receptors for direct identification of candidate compounds as agonists,
XX inverse agonists or partial agonists for use as therapeutic agents -
PS Claim 43; Page 108; 159pp; English.
XX
XX The sequence encodes a human G-protein coupled receptor (GPCR),
CC hRUP18 The endogenous and non-endogenous, constitutively activated
CC


```
XX PF 31-MAR-2000; 2000WO-US08621.
XX KW Antisense oligonucleotide; multiple target; antisense treatment;
XX KW impaired respiration; inflammation; lung disease;
XX KW pulmonary vasoconstriction; inflammation; allergic rhinitis;
XX KW acute asthma; allergy; asthma; inflammation; respiratory;
XX KW respiratory distress syndrome; pain; cystic fibrosis;
XX KW pulmonary hypertension; pulmonary vasoconstriction; emphysema;
XX KW chronic obstructive pulmonary disease; leukemia; lymphoma; carcinoma;
XX KW colon cancer; breast cancer; lung cancer; pancreatic cancer;
XX KW hepatocellular carcinoma; kidney cancer; melanoma; hepatic metastasis;
XX KW prostate cancer; ss.
XX OS Synthetic.
XX PN W09913886-A1.
XX XX
XX PD 25-MAR-1999.
XX XX
XX PF 17-SEP-1998; 98WO-US19419.
XX XX
XX PR 09-JUN-1998; 98US-0093972.
XX PR 17-SEP-1997; 97US-0059160.
XX XX
XX PA (UYEC-) UNIV EAST CAROLINA.
XX PI Nyce JW;
XX XX
XX DR WPI; 1999-229400/19.
XX XX
XX PT New antisense oligonucleotides used in treatment of, e.g. pulmonary
XX PT vasoconstriction
XX PS Disclosure; Page 37; 120pp; English.
XX XX
XX CC The specification describes antisense oligonucleotides (AA52869-X55271)
XX CC directed against at least 2 mRNAs selected from target genes, coding and
XX CC non-coding regions of RNAs corresponding to target genes, gene
XX CC initiation codons, genomic flanking regions, intron-exon borders,
XX CC 5'-end, the 3'-end and the juxta-section between coding and non-coding
XX CC regions and all segments of RNAs encoding proteins associated with one
XX CC or more diseases, conditions or mixtures. The antisense oligonucleotides
XX CC may be derived from sequences AAX5272-74. These multiple target
XX CC oligonucleotides (specifically AAX5180-271) can be used for the
XX CC antisense treatment of diseases and conditions. Typical diseases and
XX CC conditions are those associated with impaired respiration and
XX CC inflammation, including lung diseases, pulmonary vasoconstriction,
XX CC inflammation, allergic rhinitis, acute asthma, allergies, asthma, impeded
XX CC respiration, respiratory distress syndrome, pain, cystic fibrosis,
XX CC pulmonary hypertension, pulmonary vasoconstriction, emphysema, chronic
XX CC obstructive pulmonary disease (COPD), and cancers such as leukemias,
XX CC lymphomas, carcinomas e.g. colon cancer, breast cancer, lung cancer,
XX CC pancreatic cancer, hepatocellular carcinoma, kidney cancer, melanoma,
XX CC hepatic metastases, as well as all types of cancers which may metastasize
XX CC or have metastasized to the lungs, including breast and prostate cancer.
XX SQ
XX SQ Sequence 114955 BP; 6071 A; 29417 C; 36712 G; 21328 T; 21427 other;

Query Match 3.5%; Score 48.6; DB 20; Length 114955;
Best Local Similarity 30.5%; Pred. No. 0.64;
Matches 210; Conservative 79; Mismatches 397; Indels 2; Gaps 2;

Qy 642 gaggtgtggtccagcgtctctctcagcgtccctgagcgtccctgagcagcagcagcagc 701
Db 104986 GCGGCGCGCCCGCCVGNHNNNSCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 104927
Qy 702 ctgagcccaactatgcaggtgagccatcagtcgcatcagagcgtcagcgtcgtgtgagc 761
Db 104926 NSCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 104867
Qy 762 ggaacagaggtgtccctgtcaggtgaggtgaaactgttgaagtcattcacaagctcgtgacg 821
Db 104866 CVGCGGNHNNNSCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 104807

RESULT 15
AAX53491/c
ID AAX53491 standard; DNA; 114955 BP.
AC AAX53491;
XX
XX 05-JUL-1999 (first entry)
XX DT
XX DE Human adenosine A1 receptor antisense oligonucleotide fragment.
XX XX
```


THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 11, 2002, 16:04:08 ; Search time 814.02 Seconds
(without alignments)
3475.557 Million cell updates/sec

Title: US-09-820-005-3_COPY_1_3300

Perfect score: 3300

Sequence: 1 Tactaaatacaaaattag.....gtgctaggattacaagcgtg 3300

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N_Geneseq_1101:*

- 1: /SIDSB/gcgdata/geneseq/geneseq/NA1981.DAT:*
- 2: /SIDSB/gcgdata/geneseq/geneseq/NA1981.DAT:*
- 3: /SIDSB/gcgdata/geneseq/geneseq/NA1982.DAT:*
- 4: /SIDSB/gcgdata/geneseq/geneseq/NA1983.DAT:*
- 5: /SIDSB/gcgdata/geneseq/geneseq/NA1984.DAT:*
- 6: /SIDSB/gcgdata/geneseq/geneseq/NA1985.DAT:*
- 7: /SIDSB/gcgdata/geneseq/geneseq/NA1986.DAT:*
- 8: /SIDSB/gcgdata/geneseq/geneseq/NA1987.DAT:*
- 9: /SIDSB/gcgdata/geneseq/geneseq/NA1988.DAT:*
- 10: /SIDSB/gcgdata/geneseq/geneseq/NA1989.DAT:*
- 11: /SIDSB/gcgdata/geneseq/geneseq/NA1990.DAT:*
- 12: /SIDSB/gcgdata/geneseq/geneseq/NA1991.DAT:*
- 13: /SIDSB/gcgdata/geneseq/geneseq/NA1992.DAT:*
- 14: /SIDSB/gcgdata/geneseq/geneseq/NA1993.DAT:*
- 15: /SIDSB/gcgdata/geneseq/geneseq/NA1994.DAT:*
- 16: /SIDSB/gcgdata/geneseq/geneseq/NA1995.DAT:*
- 17: /SIDSB/gcgdata/geneseq/geneseq/NA1996.DAT:*
- 18: /SIDSB/gcgdata/geneseq/geneseq/NA1997.DAT:*
- 19: /SIDSB/gcgdata/geneseq/geneseq/NA1998.DAT:*
- 20: /SIDSB/gcgdata/geneseq/geneseq/NA1999.DAT:*
- 21: /SIDSB/gcgdata/geneseq/geneseq/NA2000.DAT:*
- 22: /SIDSB/gcgdata/geneseq/geneseq/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
C 1	358.4	10.9	32152	22	AA157791 Human colorectal c
C 2	356.4	10.8	32204	22	AA157790 Human colorectal c
C 3	339.8	10.3	10503	22	AA163970 Human polynucleoti
C 4	337.4	10.2	39198	22	AAF58067 Human polyamine-mo
C 5	333.2	10.1	13864	22	AAF97861 Human neuroblastom
C 6	332.6	10.1	20444	22	AA162640 Human breast or ov
C 7	329.8	10.0	1334	21	AAAB1655 N. meningitidis pa
C 8	326.6	9.9	12162	21	AAAC66548 Human kinesin-like
C 9	324	9.8	7588	22	AA162714 Human breast or ov
C 10	323.8	9.8	21721	20	AAH83427 Human lipolysis st
C 11	323.8	9.8	22976	20	AAH83426 Genomic region con

ALIGNMENTS

RESULT 1
AA157791/c
ID AA157791 standard; DNA: 32152 BP.

XX AC AA157791;

XX DF 19-OCT-2001 (first entry)

XX DE Human colorectal cancer antigen coding sequence SEQ ID NO: 328.

XX KW Human; colorectal cancer; colorectal cancer antigen; gene therapy; ds.

XX OS Homo sapiens.

XX PN WO20015350-A1.

XX PD 02-AUG-2001.

XX PF 17-JAN-2001; 2001WO-US01350.

XX PR 31-JAN-2000; 2000US-0179065.

XX PR 04-FEB-2000; 2000US-0180628.

XX PR 24-FEB-2000; 2000US-0184664.

XX PR 02-MAR-2000; 2000US-0186350.

XX PR 16-MAR-2000; 2000US-0189874.

XX PR 17-MAR-2000; 2000US-0190076.

XX PR 18-APR-2000; 2000US-0198123.

XX PR 19-MAY-2000; 2000US-0205515.

XX PR 07-JUN-2000; 2000US-0209467.

XX PR 28-JUN-2000; 2000US-0214886.

Human lipolysis st
Human leptin fragm
Human PG-3 gene.
Human G713 3'-end
Human polyamine-mo
Nucleic acid sequ
Human glycosyl sul
Human nibrin DNA.
Nucleotide sequenc
Retinoblastoma bin
Human secreted pro
Human PKD1 gene.
Human polycystic k
Human PKD1 locus b
Human prothrombin
Human polynucleoti
Human neuroblastom
Human cDNA sequenc
Human kidney relat
Human kidney relat
Human heparin cofa
Human neuroblastom
Human colorectal c
Human colorectal c
Human kidney amino
Human PACAP genom
Human neuroblastom
Human cDNA sequenc
Human colorectal c
Human SHOX (short
Human cDNA clone f
Human fascin DNA f
Human fascin DNA f
Human cDNA sequenc


```
Query Match      10.9%; Score 358.4; DB 22; Length 32152;
Best Local Similarity 77.2%; Pred. No. 2.4e-82;
Matches 460; Conservative 0; Mismatches 111; Indels 25; Gaps 1;

QY 2730 accttctctttttgtttatattttttgtttgttttgagacagagcttctgattgtgtccag 2789
DB 6938 ACCATTCCACAATTTTFTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 6879

QY 2790 qcacagatgcagtggaacgatctcagctcactcgaacctccgcctccaggttcaactgg 2849
DB 6878 GCTGGAGTGCAGTGGCCCAATCTTGGCTCACTGCAACCTCTGCTCTGTGATTCACGAA 6819

QY 2850 ttctctgcctcagctccagcagtagttggattacaggtggccaccaccacgcctggct 2909
DB 6818 TTATCCGCTCAGCCCTCAGTAGTGGATTACAGGCGCCGCCACGACGCTGGCTGGCT 6759

QY 2910 aattttgtatttttagtagagacagaggtttaccatgctggcagactggtctcaact 2969
DB 6758 AATTTTGTATTTTGTAGTAAACAGAGATTTCACCATGTTGGCCAGGCTGGCTCAAAACA 6699

QY 2970 cctgacctcaagtgcaccaccactggcctcccaagtgcctgggattacaggcacatgag 3029
DB 6698 CCTGATCTCAGGTGATCCACTGGCTTGGCCTCCCAAGTCTGGAATTACAGGCGTGAG 6639

QY 3030 ccacccgtcctggcctgtttttt-----gaaatgaggtc 3064
DB 6638 CCACAGTGCCCGCCCTTTTTTTTTTTTTTTTGTAGATGCTCGCTCCATCACCCAGGC 6579

QY 3065 tgaagtgcagtggtgcatacatagttcaactgcagcctcaactccagggcccaagtgatc 3124
DB 6578 TGGAGTGCAGTGTGCAATCTCCACTCACTGCAACCTCTGCTGGGTTCGAGTGATT 6519

QY 3125 ctctgcctcagcccttgatgtagctaggctacagcgacacacacacatgcctggctag 3184
DB 6518 CTCCTGCCTCAGCTTCAACAGTAGCTGGGACTACAGGCACATGCCACATGCCCGAGCTAA 6459

QY 3185 tttttaaattttgtgagagatgaggtttcactatgtttccaggctaaatttgaactcc 3244
DB 6458 TTTTGTGAATTTAGTAGAGACGGGTCTCAACACATTTGCCAGGCTGGTCTGAATCC 6399

QY 3245 tcggcttaagaacctctggtctcagctccacacagtgtagattacaagcgtg 3300
DB 6398 TGACCTCAAGTAATCCACCGCCCTCAGCCTCCCAAGTGTGGGATTACAGGTGTG 6343

RESULT 2
AAI57790/c
ID AAI57790 standard; DNA; 32204 BP.
XX
AC AAI57790;
XX
DT 19-OCT-2001 (first entry)
XX
DE Human colorectal cancer antigen coding sequence SEQ ID NO: 327.
XX
KW Human; colorectal cancer; colorectal cancer antigen; gene therapy; ds.
XX
OS Homo sapiens.
XX
PN WO200155350-A1.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01350.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225288.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226688.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233085.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
```


[illegible]

PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226686.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 01-SEP-2000; 2000US-0229345.
PR 03-SEP-2000; 2000US-0229509.
PR 03-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 13-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239335.
PR 20-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 17-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249285.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
(HUMA-) HUMAN GENOME SCI INC.
Rosen CA, Barash SC, Ruben SM;
WPI; 2001-488785/53.
New isolated nucleic acids and polypeptides, useful for diagnosing,
treating and/or preventing human diseases and disorders -
Disclosure; SEQ ID NO: 290; 520pp + Sequence Listing; English.
The present invention provides the protein and coding sequences of a
number of ovarian and breast antigens. These are shown in
AA162467-AA162572 and AA42240-AA42345. The sequences can be used in the
diagnosis, prevention and treatment of breast and ovarian cancers, and
their metastases. The present sequence is a genomic sequence of the

CC invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 XX
 SQ Sequence 20444 BP; 6413 A; 3836 C; 4185 G; 6010 T; 0 other;

Query Match 10.1%; Score 332.6; DB 22; Length 20444;
 Best Local Similarity 77.8%; Pred. No. 9.7e-76;
 Matches 444; Conservative 0; Mismatches 114; Indels 13; Gaps

QY 2742 ttgtgtattttttgtgttttgagacagagcttctgtctgttcgccaggccagagtgag 2801
 Db 5433 TGTCTTTTTTTTTTTTGGAGACAGAGTCTGTCTGTGCCCAGCTGGAGTGCAA 5374
 QY 2802 tggcacgactcagctcaactgcaacctccgccttccaggttcaactgggtctctcctgcctc 2861
 Db 5373 TGGCGGATCCGAGCTCAGTCACTGCAACCACTGCGCTACTGGGTCCAAGGAATTCGCGTCTC 5314
 QY 2862 agctccagagtagttgggattacagagtgccaccaccacgcctggctaa-tttttgtat 2920
 Db 5313 AGCTCCCAAGCAGCTGGGATTACAGGAGCGTGCCACCACTCGGCTAAATTTTGTGAT 5254
 QY 2921 ttttagtagacagaggttttccacctgtttggccagactggttcaaaactcctgacctcaa 2980
 Db 5253 TTTTGTAGTAGAGACGAGGTTTTCACCATGTTGGCCAGGCTGTCTCGAACCTCTGA-CTCAA 5195
 QY 2981 gtatccaccacactgggctcccaagtgctgggattacagcatagagccaccgtgctc 3040
 Db 5194 GTGATCTGCCCCACTTGGCCCTCCCAAGTCTGGGATTACAGGTGTGCGCCACCAATGCT 5135
 QY 3041 ggctgtgttttt-----tgaaatgaggtctggagtgcggtggtgcgcatcatagt 3089
 Db 5134 GGCCTTGTGTTTCTTTGAGCAGGGTTTCACAGGCTGGAGTGCGAGGGGCACAATCTTAGC 5075
 QY 3090 tcactgcagcctcaaacctccaggcccaagtgtatctctctgcctcagcccttgagtagc 3149
 Db 5074 TTATGCAATCTCCGGCTCCGGGTTCAGCGAATTTTGTGCTCAGCCACCACTAGC 5015
 QY 3150 tgagggtcacagcgccacacacacatgcctggctagtttttaaaaattttgtgagatgag 3209
 Db 5014 TGGGATTACAGGCGCCCTGCCACCATGCTTGCTTAATTTTGTATTTTATGATAGACAGG 4955
 QY 3210 gtttcactatgttgcaggctaatcttgaacctcctcggttaagcaacccctctggtctc 3269
 Db 4954 GTTTTGCCATGTTGGCAGGCTGGTCTTGAACCTCTGGCTCAAGTGATCCACCACCTC 4895
 QY 3270 agctcccaagctgctaggattacaagcgtg 3300
 Db 4894 AGCTCCCAAGTCTGGGATTACAGGCGTG 4864

RESULT 7
 AAA81655/c
 ID AAA81655 standard; DNA; 1334 BP.
 XX
 XX AAA81655;
 XX
 DT 04-DEC-2000 (first entry)
 XX
 DE N. meningitidis partial DNA sequence gnm_202 SEQ ID NO:202.
 XX
 DE
 KW Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenic;
 KW antigen; vaccine; diagnosis; infection; antibacterial; identification;
 KW Meningococcus B; MenB; ds.
 XX
 XX Neisseria meningitidis.
 OS
 XX
 XX W0200022430-A2.
 PN
 XX
 XX 20-APR-2000.
 PD
 XX

PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226688.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249254.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 11-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
(HUMA-) HUMAN GENOME SCI INC.
Rosen CA, Barash SC, Ruben SM;
WPI; 2001-488785/53.
New isolated nucleic acids and polypeptides, useful for diagnosing, treating and/or preventing human diseases and disorders -
Disclosure; SEQ ID NO: 364; 520pp + Sequence Listing; English.
The present invention provides the protein and coding sequences of a number of ovarian and breast antigens. These are shown in AAI62467-AAI62572 and AAM42240-AAM42345. The sequences can be used in the diagnosis, prevention and treatment of breast and ovarian cancers, and their metastases. The present sequence is a genomic sequence of the invention.
Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO


```
FT exon /note= "potential 5' regulatory region"  
FT 2001..2356  
FT /*tag= b  
FT /number= 1  
FT intron 2357..3539  
FT /*tag= c  
FT /number= 1  
FT exon 3540..3884  
FT /*tag= d  
FT /number= 2  
FT intron 3885..12162  
FT /*tag= e  
FT /number= 2  
FT exon 12163..12282  
FT /*tag= f  
FT /number= 3  
FT intron 12283..15143  
FT /*tag= g  
FT /number= 3  
FT exon 15144..15200  
FT /*tag= h  
FT /number= 4  
FT intron 15201..15764  
FT /*tag= i  
FT /number= 4  
FT exon 15765..15911  
FT /*tag= j  
FT /number= 5  
FT intron 15912..19578  
FT /*tag= k  
FT /number= 5  
FT exon 19579..19752  
FT /*tag= l  
FT /number= 6  
FT intron 19753..19898  
FT /*tag= m  
FT /number= 6  
FT exon 19899..19958  
FT /*tag= n  
FT /number= 7  
FT intron 19959..20055  
FT /*tag= o  
FT /number= 7  
FT exon 20056..20187  
FT /*tag= p  
FT /number= 8  
FT intron 20188..20328  
FT /*tag= q  
FT /number= 8  
FT exon 20329..20957  
FT /*tag= r  
FT /number= 9  
FT intron 20958..21046  
FT /*tag= s  
FT /number= 9  
FT exon 21047..21187  
FT /*tag= t  
FT /number= 10  
FT variation replace(818,G)  
FT /*tag= u  
FT /standard_name= "single nucleotide polymorphism"  
FT /note= "marker 17-2-297 (A31)"  
FT variation replace(1243,T)  
FT /*tag= v  
FT /frequency= 0.15  
FT /standard_name= "single nucleotide polymorphism"  
FT /note= "marker 9-19-148 (A10)"  
FT variation replace(1374,A)  
FT /*tag= w  
FT /standard_name= "single nucleotide polymorphism"  
FT /note= "marker 9-19-256 (A32)"  
FT variation replace(1401,T)  
FT /*tag= x
```

```
FT /*frequency= 0.12  
FT /standard_name= "single nucleotide polymorphism"  
FT /note= "marker 9-19-307 (A11)"  
FT delete(1535)  
FT /*tag= y  
FT /standard_name= "single nucleotide polymorphism"  
FT /note= "marker 9-19-442 (A12)"  
FT variation replace(1788, A)  
FT /*tag= z  
FT /standard_name= "single nucleotide polymorphism"  
FT /note= "marker 9-20-187 (A13)"  
FT variation replace(2391, G)  
FT /*tag= aa  
FT /frequency= 0.24  
FT /standard_name= "single nucleotide polymorphism"  
FT /note= "marker 9-1-308 (A14)"  
FT variation replace(3778,T)  
FT /*tag= ab  
FT /frequency= 0.29  
FT /standard_name= "single nucleotide polymorphism"  
FT /note= "marker 9-3-324 (A15)"  
FT variation replace(4498,T)  
FT /*tag= ac  
FT /frequency= 0.22  
FT /standard_name= "single nucleotide polymorphism"  
FT /note= "marker 99-14419-424 (A16)"  
FT variation replace(15007,A)  
FT /*tag= ad  
FT /frequency= 0.35  
FT /standard_name= "single nucleotide polymorphism"  
FT /note= "marker 9-24-260 (A17)"  
FT variation replace(15233,A)  
FT /*tag= ae  
FT /frequency= 0.15  
FT /standard_name= "single nucleotide polymorphism"  
FT /note= "marker 9-24-486 (A18)"  
FT variation replace(15826,T)  
FT /*tag= af  
FT /frequency= 0.01  
FT /standard_name= "single nucleotide polymorphism"  
FT /note= "marker 9-6-187 (A19)"  
FT variation replace(19567,G)  
FT /*tag= ag  
FT /frequency= 0.35  
FT /standard_name= "single nucleotide polymorphism"  
FT /note= "marker 9-7-148 (A20)"  
FT variation replace(19744,A)  
FT /*tag= ah  
FT /frequency= 0.14  
FT /standard_name= "single nucleotide polymorphism"  
FT /note= "marker 9-7-325 (A21)"  
FT variation replace(19786,C)  
FT /*tag= ai  
FT /standard_name= "single nucleotide polymorphism"  
FT /note= "marker 9-7-367 (A22)"  
FT variation replace(20158,G)  
FT /*tag= aj  
FT /frequency= 0.05  
FT /standard_name= "single nucleotide polymorphism"  
FT /note= "marker 9-9-246 (A23)"  
FT variation delete(20595..20597)  
FT /*tag= ak  
FT /frequency= 0.26  
FT /note= "marker LSRX9-BM (17-1-240) (A24)"  
FT variation replace(21108,G)  
FT /*tag= al  
FT /standard_name= "single nucleotide polymorphism"  
FT /note= "marker LSRX10-BM (A25)"  
FT variation replace(606,T)  
FT /*tag= am  
FT /standard_name= "single nucleotide polymorphism"  
FT /note= "marker A'1"  
FT insert(5141,G)  
FT variation
```



```
FT primer_bind 1602..1621
FT /*tag= w
FT complement(1616..1635)
FT primer_bind /*tag= x
FT 1769..1787
FT misc_binding /*tag= y
FT allele replace(1788,C)
FT /*tag= z
FT misc_binding complement(1789..1807)
FT /*tag= aa
FT CDS 2001..2190
FT /*tag= ab
FT /product= "LSR"
FT exon /note= "this sequence contains introns"
FT 2001..2356
FT /*tag= ac
FT /number= 1
FT primer_bind 2036..2053
FT /*tag= ad
FT primer_bind 2062..2081
FT /*tag= ae
FT primer_bind complement(2074..2093)
FT /*tag= af
FT primer_bind 2084..2102
FT /*tag= ag
FT misc_binding 2372..2390
FT /*tag= ah
FT allele replace(2391,C)
FT /*tag= ai
FT misc_binding complement(2392..2410)
FT /*tag= aj
FT primer_bind complement(2563..2580)
FT /*tag= ak
FT primer_bind complement(2470..2489)
FT /*tag= al
FT primer_bind complement(2483..2500)
FT /*tag= am
FT primer_bind 3455..3474
FT /*tag= an
FT exon 3540..3884
FT /*tag= ao
FT misc_binding /number= 2
FT 3759..3777
FT /*tag= ap
FT primer_bind 3775..3792
FT /*tag= aq
FT allele replace(3778,T)
FT /*tag= ar
FT misc_binding complement(3779..3797)
FT /*tag= as
FT primer_bind complement(3882..3901)
FT /*tag= at
FT primer_bind complement(4336..4356)
FT /*tag= au
FT primer_bind 4444..4463
FT /*tag= av
FT allele replace(4498,G)
FT /*tag= aw
FT primer_bind complement(4902..4920)
FT /*tag= ax
FT misc_binding 4979..4997
FT /*tag= ay
FT misc_binding complement(4999..5017)
FT /*tag= az
FT misc_binding 5122..5140
FT /*tag= ba
FT misc_binding complement(5142..5160)
FT /*tag= bb
FT primer_bind 6638..6655
FT /*tag= bc
FT primer_bind complement(7072..7089)
FT /*tag= bd
FT misc_binding 7409..7427
```

```
FT misc_binding /*tag= be
FT complement(7429..7447)
FT /*tag= bf
FT 7995..8012
FT primer_bind /*tag= bg
FT misc_binding 8375..7393
FT /*tag= bh
FT allele replace(8394,C)
FT /*tag= bi
FT misc_binding complement(8395..9413)
FT /*tag= bj
FT primer_bind complement(8576..8593)
FT /*tag= bk
FT misc_binding 8685..8703
FT /*tag= bl
FT allele replace(8704,T)
FT /*tag= bm
FT misc_binding complement(8705..8723)
FT /*tag= bn
FT misc_binding 9009..9027
FT /*tag= bo
FT allele replace(9028,G)
FT /*tag= bp
FT misc_binding complement(9029..9047)
FT /*tag= bq
FT primer_bind 9622..9639
FT /*tag= br
FT misc_binding 9931..9949
FT /*tag= bs
FT misc_binding complement(9951..9969)
FT /*tag= bt
FT misc_binding 9958..9976
FT /*tag= bu
FT primer_bind 9984..9981
FT /*tag= bv
FT allele replace(9977,T)
FT /*tag= bw
FT misc_binding complement(9978..9996)
FT /*tag= bx
FT misc_binding 10002..10020
FT /*tag= by
FT allele replace(10021,G)
FT /*tag= bz
FT misc_binding complement(10022..10040)
FT /*tag= ca
FT primer_bind complement(10023..10040)
FT /*tag= cb
FT primer_bind 10492..10512
FT /*tag= cc
FT primer_bind complement(10546..10563)
FT /*tag= cd
FT misc_binding 11857..11877
FT /*tag= ce
FT primer_bind complement(10996..11015)
FT /*tag= cf
FT allele replace(11878,T)
FT /*tag= cg
FT misc_binding complement(11879..11897)
FT /*tag= ch
FT primer_bind 11972..11990
FT /*tag= ci
FT primer_bind 12005..12023
FT /*tag= cj
FT exon 12163..12282
FT /*tag= ck
FT /number= 3
FT primer_bind complement(12417..12436)
FT /*tag= cl
```

Query Match 9.88; Score 323.8; DB 22; Length 23187;
Best Local Similarity 75.58; Pred. No. 2e-73;
Matches 457; Conservative 0; Mismatches 112; Indels 36; Gaps 3;

2792	QY	cagagtgaagtggcacgacatctcaagctcaactgcacactccgacctccaggttccaaactggtt	2851
1859	Db	TGGAGTGCAGTGGGCAACAATCTCGGCTCACTGCAACCTCCGCTCCAGGGCTCAAGTGAT	1800
2852	QY	ctccctgcctcgaagctccagagtagttggattacaggtgccccaccacacgacctggcctaa	2911
1799	Db	CTCCTGCCCTCAACCTCCCAAGTAGCTGGGATTACAGGCACGCACACCTTGGCTAA	1740
2912	QY	tttttgatatttttagtagacacaggtttccaccatgtttggccagacagtgctccaaactcc	2971
1739	Db	TTTTTGTATTTTAAACGACAGACGGGGTTTCCACCATGTTGGCCAGGCTGGCTTGGACTCC	1680
2972	QY	tgacctcaagtgtatccacccacctgggacctcccaagtgctgggattacagggcatagacc	3031
1679	Db	TGACCTTCAGGTGATCCACCCGGCTCCGGCTCCCAAAGTACTAGGATTACAGGCGTAGCT	1620
3032	QY	accgtgcttgacctgttttttgaagtgaagctctggagtgagtggtgcgatacataagttc	3091
1619	Db	ACTTACTGTGCTCGCCCTATT---TTTTTATTTTTTGTGTTTGTATGTTTATTTCTC	1563
3092	QY	actgagcctcaacctccaggcccaagtgatctctctgcctcagcccttgagtagctg	3151
1562	Db	ACTGTAGCTCTAGCCTCCCTGGCTCAA--GATCTTTCGCCCTCAGCCTCCGAGTAACTG	1505
3152	QY	gggctacaggcgccacacacatgcctggctagttttttaaatttttggagatgaggt	3211
1504	Db	GGACACAGGTGCATGCCACCATGCTAGCTAAATTTTAAATTTTAAATAGAGATAGAT	1445
3212	QY	ttaactatgtgtccaggctaaactctgaaactcctcggtctgaagcaaacctctgctcag	3271
1444	Db	CTCACTGTGTGCCACAGGTGGTCTTGAATTCCTGGCCTCAGGTAATCTCTCCATCAG	1385
3272	QY	ctctccacagtgctaggattacaagcgtg	3300
1384	Db	CTTCCCAAAGTACTGGCATTTACAGCGCTG	1356

Search completed: January 11, 2002, 16:23:10
Job time: 26142 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 11, 2002, 11:34:22 ; Search time 8779.43 Seconds
(without alignments)
4039.108 Million cell updates/sec

Title: US-09-820-005-3_COPY_1_3300

Perfect score: 3300

Sequence: 1 tactaaaatacaaaattag.....gtgctagattacaagcgtg 3300

Scoring table:

IDENTITY_NUC

Gapop 10_0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: em_estfun:*

2: em_esthum:*

3: em_estin:*

4: em_estom:*

5: em_estpl:*

6: em_estba:*

7: em_estro:*

8: em_estov:*

9: em_htc:*

10: gb_estl:*

11: gb_est2:*

12: gb_htc:*

13: gb_gss:*

14: em_gss_fun:*

15: em_gss_hum:*

16: em_gss_inv:*

17: em_gss_pln:*

18: em_gss_pro:*

19: em_gss_rod:*

20: em_gss_vrt:*

21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	446.6	13.5	475	11	BF902291 IL2-MT017
C 2	371	11.2	371	11	RI021138 IL2-MT017
C 3	338.8	10.3	344	13	AQ069145 HS-2253_B
C 4	325.4	9.9	877	13	AQ739838 HS-5505_A
C 5	320	9.7	642	10	AL597806 DKFZP3130
C 6	317.4	9.6	679	11	BF346320 602018525
C 7	315.8	9.6	595	11	BF816106 MK2-C1012
C 8	309.8	9.4	603	10	AL138254 DKFZP547E
C 9	307	9.3	5528	13	AB012089 Homo sapi
C 10	305.6	9.3	603	10	AL138254
C 11	305.4	9.3	753	11	BF542878
C 12	304.4	9.2	963	11	BG677697 602625376

C 13	303.4	9.2	702	11	BF578427
C 14	303.4	9.2	856	13	AQ891549
C 15	303.2	9.2	660	13	AQ381918
C 16	297.8	9.0	842	13	AQ748766
C 17	295.6	9.0	691	11	BG392240
C 18	295.6	9.0	1875	12	BC009270
C 19	294.2	8.9	570	13	AZ537703
C 20	294	8.9	658	13	AQ393450
C 21	292.6	8.9	487	10	AI679002
C 22	292.2	8.9	588	13	AQ421340
C 23	291.8	8.8	746	11	BG431326
C 24	291.8	8.8	5556	13	AQ839814
C 25	289.2	8.8	605	11	BF872337
C 26	289	8.8	1345	10	AV762220
C 27	288.6	8.7	767	10	AV700498
C 28	288.2	8.7	684	11	BE908602
C 29	287.6	8.7	683	10	AV760485
C 30	287	8.7	626	10	AV760760
C 31	286.6	8.7	985	10	AL044339
C 32	286.2	8.7	834	11	BG530074
C 33	286.2	8.7	839	10	AV700988
C 34	285.8	8.7	858	10	AU119532
C 35	284.2	8.6	866	13	AQ894479
C 36	283.2	8.6	689	13	AQ318783
C 37	282.4	8.6	814	13	AF102051
C 38	281.4	8.5	523	10	AI054414
C 39	281.4	8.5	611	10	AV763492
C 40	281	8.5	644	11	BG426955
C 41	281	8.5	658	13	AQ357800
C 42	280.6	8.5	626	10	AV760760
C 43	279.8	8.5	792	11	BI225234
C 44	279.4	8.5	601	10	AV760723
C 45	279.4	8.5	618	10	AV762702

ALIGNMENTS

RESULT 1
BF902291/c

LOCUS BF902291 475 bp mRNA EST 18-JAN-2001
DEFINITION IL2-MT0179-121200-290-C04 MT0179 Homo sapiens cDNA, mRNA sequence.

ACCESSION BF902291

VERSION BF902291.1 GI:12293750

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 475)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS

Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R., Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H., Brunstein, A., deOliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.

Shotgun sequencing of the human transcriptome with ORF expressed

sequence tags

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

MEDLINE 20202663

COMMENT

Contact: Simpson A.J.G.

Laboratory of Cancer Genetics

Ludwig Institute for Cancer Research

Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,

Brazil

Tel: +55-11-2704922

Fax: +55-11-2707001

Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL

(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=IL2&tl2=IL2-MT0179-

121200-290-C04&t3=2000-12-12&t4=1)

Seq primer: puc 18 forward

sapiens genomic clone Plate=2253 Col=19 Row=N, DNA sequence.

ACCESSION AQ069145
 VERSION AQ069145.1 GI:3384344
 KEYWORDS GSS.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 344)
 AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.
 TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
 JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
 MEDLINE 99380589
 COMMENT Contact: Mahairas GG, Wallace JC, Hood L
 High Throughput Sequencing Center
 University of Washington
 401 Queen Anne Avenue North, Seattle, WA 98109, USA
 Tel: (206) 616-3618
 Fax: (206) 616-3887
 Email: jwallace@u.washington.edu
 Sequence Tagged Connector
 Plate: 2253 row: N column: 19
 Class: BAC ends
 High quality sequence stop: 344.
 FEATURES
 source
 1. 344
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="Plate=2253 Col=19 Row=N"
 /clone_lib="CIT Approved Human Genomic Sperm Library D"
 /sex="male"
 /note="Organ: sperm; Vector: pBelobAC11; BAC Clones in E-Coli DH10B"
 BASE COUNT 67 a 105 c 82 g 88 t 2 others
 ORIGIN
 Query Match 10.3%; Score 338.8; DB 13; Length 344;
 Best Local Similarity 98.8%; Pred. No. 6.7e-31;
 Matches 340; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 Oy 2809 atctcagctcactgcaacctcgccctccaggttcaactggtctcctgctcagcctccc 2868
 Db 1 ATCTCAGCTCACTGCAACTCGGCTTCCAGGTTCAACTGGTCTCTCTGCTCAGCTCC 60
 Oy 2869 agagtgtgtggattacaggtgcccaccacccagcctggtgctaattttgtatttagta 2928
 Db 61 AGAGTAGTGGGATTACAGGTGCCACCACCACCGCTGGCTAAATTTTGTATTTAGTA 120
 Oy 2929 gagacaggtttaccattgtggccagactggtctcaaacctcctgacctcagtgatcca 2988
 Db 121 GAGACAGGGTTTCACTATGTGGCCAGACTGGTGTCAACATCTCTGACCTCAAGTGATCCA 180
 Oy 2989 cccacctggcctcccaagtgcctgggattacaggatgagccaccgctgctgacctgtt 3048
 Db 181 CCCACTGNGCCTCCCAAGTCTGGGATTACAGCATGACCCACCGTGGCTGGTGGT 240
 Oy 3049 tttttgaaatgaggtctgagtgagtgagtgagtgagtgagtgagtgagtgagtgagtg 3108
 Db 241 TTTTGAATGAGGTCTGGAGTGCAGTGGTGCATGATAGTCTCACTGCAGGCTCGACCTC 300
 Oy 3109 ccaggcccaagtgtatctcctcctcagcccccttgagtagctgg 3152
 Db 301 CCAGCCCAAGTGATCT 344
 RESULT 4
 AQ739838/c 877 bp DNA 16-JUL-1999
 LOCUS AQ739838
 DEFINITION HS_5505_A1_C09_47A RPCI-11 Human Male BAC Library Homo sapiens

genomic clone Plate=1081 Col=17 Row=E, DNA sequence.

ACCESSION AQ739838
 VERSION AQ739838.1 GI:5517360
 KEYWORDS GSS.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 877)
 AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.
 TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
 JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
 MEDLINE 99380589
 COMMENT Contact: Mahairas GG, Wallace JC, Hood L
 High Throughput Sequencing Center
 University of Washington
 401 Queen Anne Avenue North, Seattle, WA 98109, USA
 Tel: (206) 616-3618
 Fax: (206) 616-3887
 Email: jwallace@u.washington.edu
 Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm) or from Resear h Genetics (info@resgen.com). BAC end Web Server: http://www.htsc.washington.edu
 Plate: 1081 row: E column: 17
 Seq primer: T7
 Class: BAC ends
 High quality sequence stop: 877.
 FEATURES
 source
 1. 877
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="Plate=1081 Col=17 Row=E"
 /clone_lib="RPCI-11 Human Male BAC Library"
 /sex="male"
 /note="Vector: pBACE3.6; Site_1: EcoRI; Site_2: EcoRI;
 Male blood DNA was isolated from one randomly chosen donor and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBACE3.6 vector at EcoRI sites"
 BASE COUNT 251 a 181 c 242 g 203 t
 ORIGIN

Query Match 9.9%; Score 325.4; DB 13; Length 877;
 Best Local Similarity 73.8%; Pred. No. 1.4e-29;
 Matches 442; Conservative 0; Mismatches 126; Indels 31; Gaps 1;
 Oy 2733 ttctctctttttgtttattttttgtttgtttgagacagagcttctgtgtgcccagccc 2792
 Db 658 TCITTTTGTCTATTATTTGTTTGTGTTTGTGAGACAGAGTCTCACTTTGATGCCAGGCT 599
 Oy 2793 agagtgcagtgccagcatctcagctcactgcaacctccgcttccaggttcaactggttc 2852
 Db 598 GGAGTACAGTGGTGAATCTCAGCTCACTGCAACCTTCGCTCCAGGTTCAAGCAATTC 539
 Oy 2853 tcttgccctcagctccagagtagttgggattacaggtgtcccccaccacgcgctggctaat 2912
 Db 538 CCTTGCCCTCAGCCTCCCAACTAGCTGGGACTACAGGTGCTTGCCTCCAGCCAGCTAAT 479
 Oy 2913 ttttgatttttagtagagacaggggtttccacatcttgccagactggtctcaactcct 2972
 Db 478 TTTTATATTTTGTAGTAGACAGAGGGTTTTCACCATGTTGGCAGGCTGTGCTTGAACCTCT 419
 Oy 2973 gacctcaagtatccaccacccactggcctccaaagtgtgggattacaggtatcagcagcca 3032
 Db 418 GACCTCAGCGCATCGCTAGCCTTGGCTTCCCAAGTGTCTGGATTATAAGCGTGAGCCA 359


```

Qy 3192 aatttttgtagagtgagg-tttcaactatgtgtgtccagggtcaatcttgaaactctctcggtc 3250
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 587 -GTTTGTAGTAGAGTGGGATTTTTCCTGCTACACAGGCTGGTCTCGAACTCTGAGCT 645

Qy 3251 taagcaacctctgtctcagctccacagtgctagattacaagcgtg 3300
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 646 CAAGCAATCTCTCCCATCTTGCTGCTCTCAAGCGCTGGGATTATAGGCATG 695

RESULT 13
LOCUS BF678427 702 bp. mRNA EST 21-DEC-2000
DEFINITION 6020586ZF1 NIH_MGC_83 Homo sapiens cDNA clone IMAGE:4250246 5',
    mRNA sequence.
ACCESSION BF678427
VERSION BF678427.1 GI:11952322
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
    Email: cgapbs@email.nih.gov
    Tissue Procurement: CLONETECH Laboratories, Inc.
    cDNA Library Preparation: CLONETECH Laboratories, Inc.
    cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
    DNA sequencing by: Incyte Genomics, Inc.
    Clone distribution: MGC clone distribution information can be
    found through the I.M.A.G.E. Consortium/LLNL at:
    http://image.llnl.gov
Plate: LHCMI071 row: k column: 15
High quality sequence stop: 562.
Location/Qualifiers
    1..702
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /clone="IMAGE:4250246"
    /clone_lib="NIH_MGC_83"
    /lab_host="DH10B (T1 phage-resistant)"
    /note="Organ: prostate; Vector: pDNR-LIB (Clontech);
    Site_1: SfiI (ggccgctcgcc); Site_2: SfiI (ggccattatggcc
    ); 5' and 3' adaptors were used in cloning as follows: 5'
    adaptor sequence: 5'-CAGGCCATTATGGCC-3' and 3' adaptor
    sequence: 5'-ATTCTAGAGCGGCGGCGGCATG-dt(30)BN-3'
    (where B = A, C, or G and N = A, C, G, or T). Average
    insert size 1.4 kb (range 0.5-4.0 kb). 14/15 colonies
    contained inserts by PCR. This library was enriched for
    full-length clones and was constructed by Clontech
    Laboratories (Palo Alto, CA)."
BASE COUNT 223 a 170 c 210 g 99 t
ORIGIN

Query Match 9.28; Score 303.4; DB 11; Length 702;
Best Local Similarity 76.3%; Pred. No. 6e-27;
Matches 442; Conservative 0; Mismatches 121; Indels 16; Gaps 5;

Qy 2734 tctctttttgtattttttgttttgagacagatcttgatct-gttgccagagcc 2792
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 595 TTTTCTTTTCTGTTCTCTTTTCTTTTGTAGACAGAGCTCTACCCCTGGTTGCCAGTCT 526

Qy 2793 anagtcagtgagcagatctcagctcactcgaacctccgctcttcaggttcaactgggttc 2852
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 525 GGAGTGTCTATGGCGCAATCTCGGCTCACTGCAGCCTCCGCTTCCGGGTTCAAAATGATTC 466

Qy 2853 tctgctcctcagcctccagagtagttggattacaggtgccccaccacacgctcggttaat 2912
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 465 TCCTGCTCAGCCTCCCGAGTAGCGGGATTACAGGTGCCCGCCACCATGCCCCAGCTAAT 406

```

```

Qy 2913 ttttgatttttagtagagacagaggttttcaccattgttggccagactggtgtctcaactcct 2972
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 405 TTTTGTATTTTGTAGTAGAGACAGAGGTTTTCACCATGTTGACCAAGGCTGCGCTCGAACTCCT 346

Qy 2973 gacctcaagtatcaccaccacctgggctcccaaaagtgctgggattaccaggcagagcca 3032
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 345 GACCTC--GTGATCCGCCCGCTCGGCTCCCAAAGTCTGGGATTACAGGTGTGAACA 288

Qy 3033 ccgtgctcgctgctgttttttgaatagaggtctgaggt-----gcagtgggtgcg 3081
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 287 CTGTGCCCGCGCTTTTGTGTTGTTGTAGTCGTCTGCTGCTGTCGAGTGGTGG 228

Qy 3082 atcatagttcactcagctcacaacctcccgagcccaagtgatcctcctcgctcagccct 3141
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 227 ATCTCGGCTCACTCAACCTCTGCTCTTGGATTCAAGCGATTCTCCGTCGCTCAGCCCTT 168

Qy 3142 tgatagctggggtctacagggcgcaaccaccatgcctcgctagtttttaaaattttgtg 3201
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 167 CCAGTAGCTAGAGCTACATGCTGCATCTCCACGCTCGCTAAATTTTGTGTTTGTGTTA 108

Qy 3202 gagatgaggtttcactatgttgcaggctaactctgaacctcctcgcttaagcaacct 3261
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 107 GAGACG--GGTTTCAACATGTTG--CCAGGCTGCTCTCGAACTCTCGGCTCAAGTCATCTG 50

Qy 3262 ctgctcagcctccacagtcgtaggattacaagcgtg 3300
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 49 CCCATCTTGGCTTCTCTAAAGTCTGGGATTATAGGTGTG 11

RESULT 14
LOCUS AQ891549/c 856 bp DNA GSS 10-NOV-1999
DEFINITION HS_3143_AL_C12_T7C CIT Approved Human Genomic Sperm Library D Homo
    sapiens genomic clone Plate=3143 Col=23 Row=E, DNA sequence.
ACCESSION AQ891549
VERSION AQ891549.1 GI:6347739
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
    Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
    Hood,L.
TITLE Sequence-tagged connectors: A sequence approach to mapping and
    scanning the human genome
JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE 99380589
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
    High Throughput Sequencing Center
    University of Washington
    401 Queen Anne Avenue North, Seattle, WA 98109, USA
    Tel: (206) 616-3618
    Fax: (206) 616-3887
    Email: jwallace@u.washington.edu
    Clones may be purchased from Research Genetics (info@resgen.com).
    BAC end Web Server: http://www.htsc.washington.edu
    Plate: 3143 row: E column: 23
    Seq primer: T7
    Class: BAC ends
    High quality sequence stop: 856.
    Location/Qualifiers
    1..856
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /clone="Plate=3143 Col=23 Row=E"
    /clone_lib="CIT Approved Human Genomic Sperm Library D"
    /sex="male"
    /note="Organ: sperm; Vector: pBelOBAC11; BAC Clones in
    E-Coli DH10B"
BASE COUNT 239 a 179 c 233 g 200 t
ORIGIN

```

Query Match 9.2%; Score 303.4; DB 13; Length 856;
Best Local Similarity 74.8%; Pred. No. 5.4e-27;
Matches 442; Conservative 0; Mismatches 123; Indels 26; Gaps 4;

QY 2735 tctcttttttggatatttttgggttttgagacagagtgatc-tgttgcagagcca 2793
DB 648 TTTTCTGTTTTTTTCTTTTTTTTGGATGGATGTTTCTCTCTTGTGCCAGGCTG 589

QY 2794 gagtacagtggcagatcagctcactcagcctccacccctccctccaggttcactgttct 2853
DB 588 GAGTCAATGGCATATTCAGTCNCCTGCAACCTCTCCCTCNCAGGTTCAAGCGATTCT 529

QY 2854 cctgcctcagcctcagagtagttgggtattacagtgccccaccacagcctggctaatt 2913
DB 528 CTGTGCTCAGCTCCGAGTAGCTGGGATTCAGGCATGCACACACGCTTGGCTAA-T 470

QY 2914 ttgtatttttagtagacaggggtttcaccatgttggccagactggctcaaaactcctg 2973
DB 469 TTTGTATTTTGTAGTAGACGCGAGTTTCGCCATGTGTGTCAGGCTGTCTCGAACTCCG 410

QY 2974 acctcaagtgtaccaccacacccctggcctcccaagtg-c-tggattacagcgatgacca 3032
DB 409 ACTTCAGTGATTCACCCCGCTTGGCCCTCCCAAGTGCNTGGGATTACAGGCATGAGCCA 350

QY 3033 cgtgcctggcctgttttttgaatgag-----gtctggag 3069
DB 349 CTGTGCCCGGCTATTTACTATTATTTAAGAAAGGATCTCAGTCGTACCCAGGCTAGAG 290

QY 3070 tgcagtgctgcgatactagttcactgcagcctcaccctccagcccaagtgatcctct 3129
DB 289 TGCAGTGTATGATCACAGCTCACTGCAGCTCAACCTCCGGGGCTCAAGCATCTCTCC 230

QY 3130 gcttcagcccttgaagtgcgggtacagcgccacacacacacacacacacacacacac 3189
DB 229 ACCTCACCTCCCACTACCTTGGGACACACAGGCATGTACCACTGCACCTGGCTAGTTTT 170

QY 3190 aaaaatttttggagatgaggtttcactatgttgcagggtaactcttgaactcctcggc 3249
DB 169 GTGTTTTTGTAGATCAGATCTGTGTTATGCTGTAGGCTGGTCTCAAACTCTGGCC 110

QY 3250 ttagcaacccctcgtgctcagcctccacacagtgctaggattacaagcgtg 3300
DB 109 TCAAGCAGCTCTCCACATCGCGCTCCCAAAAGTGTGGGATTACAGGCATG 59

RESULT 15
AQ381918 660 bp DNA GSS 21-MAY-1999
LOCUS RPC111-135N16.TV RPCI-11 Homo sapiens genomic clone RPCI-11-135N16,
DEFINITION DNA sequence.
ACCESSION AQ381918
VERSION AQ381918.1 GI:4352941
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 660)
AUTHORS Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter
J.C.
TITLE Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready
Map Building
JOURNAL Unpublished (1997)
COMMENT Other_GSSs: RPC111-135N16.TJ
Contact: Shaying Zhao, William Nierman, Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850
Tel: 301 838 0200
Fax: 301 838 0208
Email: hbe@tigr.org

Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (<http://bacpac.med.buffalo.edu/ordering>) or from Research Genetics (inforesgen.com). BAC end search page: http://www.tigr.org/tldb/humgen/bac_end_search/bac_end_search.html
Seq primer: T7
Class: BAC ends.

FEATURES
source
1..660
/organism="Homo sapiens"
/db_xref="GDB:7551783"
/db_xref="taxon:9606"
/clone="RPCI-11-135N16"
/clone_lib="RPCI-11"
/sex="Male"
/cell_type="Lymphocytes"
/note="Vector: pBAC3.6; Site_1: EcoRI; Site_2: EcoRI;
RPC111 Human Male BAC Library"

BASE COUNT 136 a 180 c 152 g 192 t
ORIGIN

Query Match 9.2%; Score 303.2; DB 13; Length 660;
Best Local Similarity 73.1%; Pred. No. 6.6e-27;
Matches 431; Conservative 0; Mismatches 128; Indels 31; Gaps 2;

QY 2742 ttgttattttttgttttgagacagagtcctgattgtgtccacagccagagtgacg 2801
DB 45 TGTATTTTTTTTATTTTTTTTGACATGGAGTCTTGGTCTGTACCCAGGCTGGAGTGAA 104

QY 2802 tggcagcatctcagctcactcgaacctccgctccctccaggttcaactgttctcctgcctc 2861
DB 105 TGGCACATCTCGGCTCACCACACCTCCGCTCTCACATTCAGTGACTCTCTCTGCCCTC 164

QY 2862 agcctccagagtagttgggtattacaggtgcccacacacacacacacacacacacacac 2921
DB 165 AGCTTCCCGAGTAGTGGGACTATAGGCACGACACACCATATCCGGCTAAATTTGTATT 224

QY 2922 tttagtagacagaggtttaccatgttggccagactggtctcaaacctccctgaccccaag 2981
DB 225 TTTAGTAGAGATGGGGTTTCACCTCTTGGCCAGGCTGGTCTCAAACTCCTGTGCTCAGG 284

QY 2982 tgatccacccacccctggcctcccaagtgctgggattacaggcacagcagcagcagcagc 3041
DB 285 TGACTCACCCACTTGACTTCCACAGTGTGGGATTCAGGTGTGAGCCACTGGGCCCG 344

QY 3042 gctcg-----ttttttgaaatgaggt-----ctggagt 3070
DB 345 GCCAACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 404

QY 3071 gcagtgtgcgatacatattcactgcagcctcaaacctccacagcccaagtgatcctcctg 3130
DB 405 GCAGTGACGCTATTTGGGCTCACTGCAACTTCTGCTCCAGGTTCAAGCGATTCTCCTA 464

QY 3131 cctcagcccttgcagtagctgggctacagcgccacacacacacacacacacacacacacac 3190
DB 465 TCTCAGCCTCTGAGTAGTGGGATTTATAGGCGCTGCGCAACATACCCAGCTAATCTTTG 524

QY 3191 aaattttgtggagtaggtttcactactgtttgtccaggttaattcttgaactcctcgctg 3250
DB 525 CATTTTACTAGGATGGGTTTCGCCATGTGGCCAGGCTGGTCTTGAACCTCTGACCT 584

QY 3251 taagcaacctctgctcagcctcccaacacagtcagtaggattacaagcgtg 3300
DB 585 CAAGTGATCCCGCCGCTTGGCCTTCAACAGTGTGGGATTTATAGGCATG 634

Search completed: January 11, 2002, 11:35:07
Job time: 8859 sec

THIS PAGE BLANK (USPTO)
THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 11, 2002, 15:53:12 ; Search time 14958.6 Seconds
(without alignments)
3639.417 Million cell updates/sec

Title: US-09-820-005-3_COPY_1_3300

Perfect score: 3300

Sequence: 1 tactaaaaatacaaaattag.....gtgctagattacaagcgtg 3300

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb_ba.*

2: gb_htg.*

3: gb_in.*

4: gb_om.*

5: gb_ov.*

6: gb_pat.*

7: gb_ph.*

8: gb_pl.*

9: gb_pr.*

10: gb_ro.*

11: gb_sts.*

12: gb_sy.*

13: gb_un.*

14: gb_vi.*

15: em_ba.*

16: em_fun.*

17: em_hum.*

18: em_in.*

19: em_om.*

20: em_or.*

21: em_ov.*

22: em_pat.*

23: em_ph.*

24: em_pl.*

25: em_ro.*

26: em_sts.*

27: em_sy.*

28: em_un.*

29: em_vi.*

30: em_htgo_hum.*

31: em_htgo_inv.*

32: em_htgo_rod.*

33: em_htgo_hum.*

34: em_htg_inv.*

35: em_htg_rod.*

36: em_htg_other.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

8

Result No.	Score	Query Match	Length	DB	ID	Description	
1	3288.8	99.7	124526	9	AC005080	AC005080 Homo sapi	
2	3267.6	99.0	169604	9	AC083884	AC083884 Homo sapi	
3	3267.6	99.0	230552	9	AC005098	AC005098 Homo sapi	
4	3265.6	99.0	131359	9	AC004883	AC004883 Homo sapi	
5	3265.6	99.0	275197	9	AC004166	AC004166 Homo sapi	
6	3172	96.1	17302	9	AF184614	AF184614 Homo sapi	
C	2179.8	66.1	184558	2	AC092405	AC092405 Papio cyn	
8	2106.8	63.8	8119	9	HS47LL14S01	HS47LL14S01	
9	2105.8	63.8	8131	9	HS47PHOX01	HS47PHOX01	
10	1978	59.9	3074	9	AF003533	AF003533 Homo sapi	
11	1943.4	58.9	1965	9	AF184613	AF184613 Homo sapi	
12	1552.2	47.0	1569	9	HS47P41S01	HS47P41S01	
13	1263.6	38.3	7258	9	HS47P43S01	HS47P43S01	
14	1260.4	38.2	7258	9	HS47P40S01	HS47P40S01	
15	1251.6	37.9	3594	9	HSNCF1S1	HSNCF1S1	
C	16	1090.4	33.0	149830	2	AC027219	AC027219 Homo sapi
17	661.2	20.0	57245	2	AC068263	AC068263 Homo sapi	
C	18	632.2	19.2	57245	2	AC068263	AC068263 Homo sapi
19	589.2	17.9	101509	2	AC027353	AC027353 Homo sapi	
20	481.4	14.6	524	9	HSU33006	HSU33006 Human p47-p	
21	462.8	14.0	486	9	HS39NCF1	HS39NCF1	
C	22	437.4	13.3	198410	2	AP000831	AP000831 Homo sapi
C	23	416.6	12.6	186925	2	AC087503	AC087503 Homo sapi
24	413.2	12.5	140161	2	AL390725	AL390725 Homo sapi	
25	410.4	12.4	115079	2	AC088105	AC088105 Homo sapi	
C	26	407.8	12.4	165341	2	AC092037	AC092037 Homo sapi
27	406	12.3	67674	2	AC090641	AC090641 Homo sapi	
28	406	12.3	108369	17	AF124523	AF124523 Homo sapi	
C	29	406	12.3	344150	9	AF235103	AF235103 Homo sapi
C	30	401.2	12.2	218445	2	AC083959	AC083959 Homo sapi
C	31	400	12.1	177990	2	AC016385	AC016385 Homo sapi
32	394.4	12.0	177990	2	AC016385	AC016385 Homo sapi	
33	394	11.9	194938	2	AC073841	AC073841 Homo sapi	
C	34	391.8	11.9	106928	2	AC005049	AC005049 Homo sapi
C	35	391.6	11.9	155046	2	AC069408	AC069408 Homo sapi
36	390.4	11.8	148822	9	AC020904	AC020904 Homo sapi	
37	390.2	11.8	42156	2	AC011542	AC011542 Homo sapi	
38	389.6	11.8	166181	9	HS37E16	HS37E16 Human DNA	
39	387	11.7	273403	2	AC011498	AC011498 Homo sapi	
40	386.6	11.7	11954	9	AL133275	AL133275 Human DNA	
41	385.8	11.7	41572	9	AC006132	AC006132 Homo sapi	
42	383.4	11.6	64681	2	AC023215	AC023215 Homo sapi	
43	382.8	11.6	198410	2	AP000831	AP000831 Homo sapi	
C	44	380.6	11.5	190225	2	AC011191	AC011191 Homo sapi
C	45	379.6	11.5	120955	9	HUAC002310	HUAC002310 Human Chr

ALIGNMENTS

RESULT 1	AC005080	124526 bp	DNA	PRI	02-OCT-2000
LOCUS	Homo sapiens BAC clone CTA-269P13 from 7q11.2, complete sequence.				
DEFINITION	AC005080				
ACCESSION	AC005080				
VERSION	AC005080.2	GI:7770715			
KEYWORDS	HTG.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
TITLE	1 (bases 1 to 124526)				
JOURNAL	Sulston, J.E. and Waterston, R.				
MEDLINE	Toward a complete human genome sequence				
AUTHORS	99063792				
TITLE	2 (bases 1 to 124526)				
JOURNAL	Scott, K., Layman, D., Kalicki, J. and Harmon, G.				
	The sequence of Homo sapiens BAC clone CTA-269P13				
	Unpublished				

```

REFERENCE
AUTHORS
TITLE
JOURNAL
3 (bases 1 to 124526)
Waterston,R.H.
Direct Submission
Submitted (12-JUN-1998) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
4 (bases 1 to 124526)
Waterston,R.H.
Direct Submission
Submitted (12-MAY-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
5 (bases 1 to 124526)
Waterston,R.H.
Direct Submission
Submitted (12-JUN-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
6 (bases 1 to 124526)
Waterston,R.
Direct Submission
Submitted (02-OCT-2000) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On May 12, 2000 this sequence version replaced gi:3212911.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: sapiens@watson.wustl.edu
----- Summary Statistics
----- Center project name: H_RG269P13
-----

NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.

MAPPING INFORMATION:
The sequence of this clone was established as part of a mapping and
sequencing collaboration between the NHGRI Chromosome 7 Mapping
Project (Eric D. Green, Director), John D. McPherson in the
Department of Genetics (Washington University), and the Washington
University Genome Sequencing Center. For additional information
about the map position of this sequence, see
http://www.nhgri.nih.gov/DIR/GRB/CHR7, send
mailto:egreen@nhgri.nih.gov, or see http://genome.wustl.edu/gsc

SOURCE INFORMATION:
Clone CTA-269P13 is from a release of the human BAC library
CITB-HS-A. The library contains cloned DNA from human sperm. See:
Shizuya et al., Proc. Natl. Acad. Sci. USA 89:8794-7 (1992); U.-J.
Kim et al., Genomics 34:213-8 (1996). The clone is available from
Research Genetics, Inc. (http://www.resgen.com).
VECTOR: pBelOBAcl1
Selection: chloramphenicol

NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the left is RP4-754G14; the clone sequenced
to the right is RP11-396K3. Actual start of this clone is at base
position 1 of CTA-269P13; actual end is at base position 124526 of
CTA-269P13.

The clone CTA-269P13 may contain a transposon in the growth of the
clone, which is not part of the submitted sequence.

```

```

FEATURES
source
Location/Qualifiers
1..124526
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="7"
/map="7q11.2"
/clone="CTA-269P13"
/clone_lib="CITB-HS-A"
1..138
/notes="match to EST AA643257 (NID:g2568475) nr61h09.s1"
1..338
/notes="match to EST D44965 (NID:gl572440)"
42..118
/notes="similar to EST D44969 (NID:gl572444)"
complement(106..396)
/db_xref="GI:3800554"
119..483
/notes="similar to EST AW236834 (NID:g6569223) xm49f08.x1"
119..483
/notes="similar to EST AI609282 (NID:g4618449) tw83h06.x1"
119..378
/notes="similar to EST D44969 (NID:gl572444)"
146..483
/notes="similar to EST N80671 (NID:gl243372) zb03a07.s1"
164..483
/notes="similar to EST AW204636 (NID:g6504108)"
204..206
/notes="match to EST AA643257 (NID:g2568475) nr61h09.s1"
306..652
/notes="match to EST AA804480 (NID:g2875993) nz33c09.s1"
335..520
/notes="similar to EST AW177547 (NID:g6443584)"
577..666
/notes="similar to EST AW177547 (NID:g6443584)"
595..711
/rpt_family="Alu"
/rpt_family="(CAAAA)n"
712..740
repeat_region
repeat_region
741..898
/rpt_family="Alu"
899..919
/rpt_family="AT-rich"
965..1106
/rpt_family="Alu"
1107..1395
/rpt_family="Alu"
1396..1570
/rpt_family="Alu"
1571..1662
/rpt_family="L1"
1665..1817
/rpt_family="L1"
1834..1886
/rpt_family="L2"
1888..2189
/rpt_family="Alu"
2256..2546
/rpt_family="Alu"
3070..3391
/rpt_family="Alu"
3493..3780
/rpt_family="Alu"
3792..4101
/rpt_family="Alu"
4108..4232
/rpt_family="Alu"
4247..4422
/rpt_family="Alu"
4423..4729
/rpt_family="Alu"
4730..4866
/rpt_family="Alu"
4941..5099
repeat_region

```


|||||
Db 107523 AGAAGGATGAGGTTGCTATGCAATGTGAAGTATCAAGGAAGGCTTCTCGAAGAGG 107582
Qy 1621 tgacatttgagcagaagaatggagagagattatgagggagagattggaatggggaac 1680
Db 107583 TGACATTGAGCAGAGAAATGGAGAGAGATTATGGAGGAAGATGGTGAATGGGGGAAC 107642
Qy 1681 atggtcaagaccaggaataatggtcaagggggggaagatggtcaaggggagcagcagaatg 1740
Db 107643 ATGCTCAAGACCAGCAATATGGTCAAGGGGGGAAGATGTCAAGGGGAGCCAGCAATG 107702
Qy 1741 caaaggccctgagcagagagcagcttctctctccctgagctggtgtgtaacatgag 1800
Db 107703 CAAAGGCCCTGAGCAGAGCAGCTTGATTCACCCCAACCCCGTGGGCCCTGCAGG 107762
Qy 1801 cgaggaagagacaagtgtaaaccccttctctctccctgagctggtgtgtaacatgag 1860
Db 107763 CGAGCGGAAGGACAAGTGTAAACCCCTTTCTCTGCTCGAGGTGTGTGAACATGAG 107822
Qy 1861 tctgcccatgtttacacccctgcaagcctgaagagtcctcccgagaaactgaagaagaacaa 1920
Db 107823 TCTGCCCATGTTTACACCCCTGCAAGCCTGAAGAGTCCCAGAACTGAAGAAGACAA 107882
Qy 1921 agcccttctgtaccctccctgccccctgtcccgaccgcgacaaaacgcgacttcccttt 1980
Db 107883 AGCCCTTTCTGTACCCCTCCCTGCCCCCTGTCCCGACCGCGACAAAACGCACTTCCCTCTT 107942
Qy 1981 ccaagtcaattaaagcagcagcctgaagatgccagggagcactgagagccaccagatcag 2040
Db 107943 CCAGTGATATTAAGCGCAGCCTGGAAGTCCAGGAGACACTGGAGGCCACCCAGTCA 108002
Qy 2041 ggggacaccttcacccgtcacatcgccctgctgggctttgagaagccttcgtaccacgc 2100
Db 108003 GGGGACACCTTCATCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 108062
Qy 2101 cagcactatgtagtagctggtgaggggcatcccggtgggggaatacgggagggagacga 2160
Db 108063 CAGCACATGTGAGTAGCTGTGTGAGGGCATCCCGTGGGGGAATACGGGAGGACAGCA 108122
Qy 2161 cggccacccttgagtcaccagggcccaaccagctccagtgaggactaacggggaaggtct 2220
Db 108123 CGGCCACCTTGCAGTCCCAAGGGCAACCACTCCAGTCCAGTGAAGGACTAAGCGGGCAGGCT 108182
Qy 2221 tgggcaacctgctccctgctgtgagcctggtatccctctgacccctgaggaagacag 2280
Db 108183 TGGGCACCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 108242
Qy 2281 ttccttggaccctgctgggcccagcccttactgtcccgccctgtgtcccccagccag 2340
Db 108243 TTCCCTTGGACCCGCTGGGGCCCGACGCCCTTTACTGTGCCCGCTGTGTCCCGCCAGC 108302
Qy 2341 gcccagccttagcagagagctctctctctgctccctgacccctgagcagcagccagc 2400
Db 108303 GCCCTAGCCTTAGCAGAGAGTCTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 108362
Qy 2401 gctctcagctcagagcccaactctccaggaagccttccctgactagccagctatca 2460
Db 108363 GCTCTCAGGTCCGAGGCCACTCTCCAGGAAGCCTTCCCTGACTAGCCAGCTATCA 108422
Qy 2461 gagagtgccctcccaagagggagcctgaaactaaagctctctctctcccccagctgcc 2520
Db 108423 GAGAGTGGCCCTCCCAAGAGGGGCGCTGGAACATAAAGCTCTCTCTCTCCCGAGCTGCC 108482
Qy 2521 tgtagctcagtagagcttactctctccagtagggtgacacatcacaggggccaata 2580
Db 108483 TGTAGTGTCAAGTGTATCTCTCTCAGTAGGGGTGACACCATGACAGGGGCCAATA 108542
Qy 2581 gagtctcccatctgtcccaagagagctgtgacaaaacgctgctcagacacacagtcaca 2640
Db 108543 GAGTCTCCCATCTGTCCCAAGAGGCTGGACAAATGCCTGCTCAGACACACAAAGTCCA 108602
Qy 2641 ctgggtcccttaatcccataggaagggccagggaggaactacattaggaattgaagctt 2700
|||||

Db 108603 CTGGTCCCCTAATCCCATAGGAAGCCAGGAGAACTACATTTAGAAATTAAGACTT 108662
Qy 2701 gtatggaacatttagtctctatgtgccagaccttctctcttttttttttttttttttt 2760
Db 108663 GTATGGAAACATTTAGTCTCTATGTCGCCAAGACCTTTCTTTTTTTTGTATTTTTT 108722
Qy 2761 ttgagacagagctcttctctctgtgtgccagggccagagtgcaagtgcagatctcagctcac 2820
Db 108723 TTGAGACAGAGTCTTGATCTGTTGCCAGGCCAGAGTGCAAGTGCCAGTCTCAGCTCAC 108782
Qy 2821 tgaacacctccagctccaggttcaactggttctcctcctcagcagcagcagtagtggg 2880
Db 108783 TGAACCTCCGCTTCCAGGTTCAACTGGTTCTCCTCCTCAGCTCCAGAGTAGTGGG 108842
Qy 2881 attacagtgccacaccacccagcctgctgaatttttttttttttttttttttttttttt 2940
Db 108843 ATTACAGGTGCCACACCACCGCTGGCTAATTTTGTATTTTTTAGTAGACAGGGTTT 108902
Qy 2941 caccatgttgccagagctggtctcaaacctcctgaccccaagtgcacccaccacccctggcc 3000
Db 108903 CACCATGTTGSCCAGACTGGTCTCAAACTCCTGACCTCAAGTGATCCACCCACCTGGGCC 108962
Qy 3001 tccaaagtgctgggattacagggcatgagccaccgtgacctgacctgttttttttttttttt 3060
Db 108963 TCCAAAGTGTGGGATTTACAGGCATGAGCCACCGTGGCTGTGTTTTTTTGAATGA 109022
Qy 3061 ggtctgagtgagtggtgcatcatagttcaactgacgacctgaacctcccgagcccaagt 3120
Db 109023 GGTCTGAGTGACGTGGTGGCATCATAGTTCACTGCAAGCTGCGCTCCCGAGGCCCAAGT 109082
Qy 3121 gatcctcctgctcagcccttgagtagtggtgagtgagggctacagcgccacaccacccatg 3180
Db 109083 GATCCTCCTGCCTCAGCCCTTGAGTAGTGGGGCTACAGCGGCACACCCACCATGSCCTGG 109142
Qy 3181 ctagttt 3240
Db 109143 CTAGTTTTTTAAATTTTTTGTGGAGATGAGGTTTCACTATGTTGTCAGGCTTAATCTTGA 109202
Qy 3241 ctctcgcttaagcaacctctgctcagctccacagctgctagattacaagctg 3300
Db 109203 CTCTCGCTTAAAGCAACCCCTCTGCTCAGCCCTCCACAGTGTAGATTACAAGCGTG 109262

RESULT 2

AC083884 169604 bp DNA PRI 23-MAY-2001

LOCUS Homo sapiens clone RP11-813J7, complete sequence.

DEFINITION AC083884

AC083884.6 GI:14190780

AC083884.6

KEYWORDS

HTG.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

On May 23, 2001 this sequence version replaced gi:13431264.

Center project name: H_NH0813J07.

Location/Qualifiers

FEATURES

Db 96598 CCAGTGCATTTAAAGCGCAGCGCTTGGAAAGTGCCAGGGAGCACTGGAGGCCACCCAGTCA TG 96657
Qy 2041 ggggacaccttcacgcgcacatcgccctgctgagcttgaagagccttcgtaccacgc 2100
Db 96658 GGGACACCTTCATCCGTCACATCGCCCTGCTGGGCTTTGAGAAAGCGTTCGTACCCAGC 96717
Qy 2101 cagcactatgtagtagctggtgagggcattcccgctgggggaatacgggagggacagca 2160
Db 96718 CAGCACATATGTAGTAGTGGTGAGGGCATCCCGTGCGGGGATACGGGAGGCACAGCA 96777
Qy 2161 cggcacccttgagtcaccaggggccaaaccagctccagtggaggactaacggggcaggtct 2220
Db 96778 CGGCACCCCTTGCAGTCCCAAGGGCAACACAGCTCCAGTGGAGACTAACGGGGCAGGGTCT 96837
Qy 2221 tgggacacttgctccctgctttagccttgagctatcaccctctgactccctcctgaggaagacag 2280
Db 96838 TGGGACCTTGCTCCCTGCTTTTGGGCTTGGATCTACCCCTCTGATCCCTTGGGAGACAG 96897
Qy 2281 ttcccttgaccctccctgggcccacagccctttactgtcccccctgtgtcccccagcag 2340
Db 96898 TTCCCTTGGACCCGCTTGGGCCCCAGCCCTTTACTGTCCCGCCTGTGTCCCCAGCCAG 96957
Qy 2341 gccctcagccttagccaggagctctcttctgtctccctgctccatggccaggcagccacgc 2400
Db 96958 GCCCTCAGCCTTAGCCAGGAGCTCTCTTCTGCTCCCTGCTGCTGCGCAGGCAAGCCAGC 97017
Qy 2401 gctctcagctccgagcccaactcctccaggagccttccctcactagccagcactatca 2460
Db 97018 GCTCTCTAGTCTCGAGGCCCACTCTCCAGGAAGCCTTCCCTGACTAGCCACGACTATCA 97077
Qy 2461 gagagtgccctcccaagagggagccttggaaactaaagctctctctcccccagctgcc 2520
Db 97078 GAGAGTGGCCCTCCCAAGAGGAGCGCTGGAAACTAAAGCTCTCTCTCCCCAGCTGCC 97137
Qy 2521 tgtagtgctcagtagagcttctctcctccagtgagtgagtgagcagcagggcccaata 2580
Db 97138 TGTAGTCTCAGTTAGAGTCTTATCCTCTCCAGTAGGCTGACACCATGACAGGGGCCAATA 97197
Qy 2581 gaggctcccatctgtcccaagagggctgagcaaatgcctgctcagacacacagtcaca 2640
Db 97198 GAGTCTCCCATCTGTCCCAAGAGGCTGGACAAATGCTGCTCAGACACACAAGTCCA 97257
Qy 2641 ctgggtcccttaatacccataggaagggccaggaggaactacattaggaattgaagctt 2700
Db 97258 CTGGGTCCCTTAATCCCATAGGAAGGCCAGGAGGAACACTACATTTAGGAATTTGAAGCTT 97317
Qy 2701 gtaggaacatttagctctatgtcccaagaccttctctttttttttttttttttttttgt 2760
Db 97318 GTATGGAACTTTAGTCTATGTGCCAAGACCTTCTCTTTTGTATTTTGTATTTTGTGTTT 97377
Qy 2761 ttgagacagagcttctgtctgttgcacagggcagagtgagtgacagatctcagctcac 2820
Db 97378 TTGACACAGAGCTCTGTATCTGTTCGCCAGGCCACAGTGCACTGGCAGCATCTCAGCTCAC 97437
Qy 2821 tgcacacctccgcttccaggttccactggttctcctcctcagcctccagtagtggtggg 2880
Db 97438 TGCAACCTCCGCTTCCAGGTCAACTGGTTCTCTGCTCCTCAGCCTCCAGAGTAGTGGG 97497
Qy 2881 attacaggtgcccaccaccagcctggctaatttttatttttttttttttttttttttttt 2940
Db 97498 ATTACAGTGGCCCAACCAACCGCTGGCTAATTTTGTATTTTGTATGAGACAGGGGTTT 97557
Qy 2941 caccatgttgccagagctggtctcaactcctcagctcaagtgcacccacccacctgggcc 3000
Db 97558 CACCATGTTGGCCAGAGTGGTCTCAAACTCTGACCTCAAGTGATCCACCCACCTGGGCC 97617
Qy 3001 tcccaagtgctgggattacagggcatgagccaccgtgcttggcctgttttttttttttttttt 3060
Db 97618 TCCCAAGTGTGGGATTACAGGATGAGCCACCGCTGCTGGCCTGTTTGTGAAATGA 97677
Qy 3061 ggtctgagtgagtggtgagatcatagttcactgcagctcaacctccagccagcccaagt 3120
Db 97678 GGTCTGGAGTGCAGTGGTGGCATCATAGTTTCACTGCAAGCCTTCGACCTCCCAAGGCCCAAGT 97737

Qy 3121 gatcctcctgctcagcccttgagtagtgagggtacagggcacaccaccatgcctgg 3180
Db 97738 GATCCTCTGCTCAGCCCTTGAGTACCTGGGCTACAGCGCACACCACCATGCGCTGG 97797
Qy 3181 ctagtttttaaatttttggagatgaggtttcactatgttgcagggttaattttgaa 3240
Db 97798 CTAGTTTTAAATTTTGTGGAGATGAGGTTTTCATGTATGTTGTCCAGGCTAATCTTGAA 97857
Qy 3241 ctctcggcttaagaacacctctggtctcagcctccacagtcctcagtaggattacaagcgtg 3300
Db 97858 CTCCTGCGCTTAAAGCAACCCCTCTGCTCAGCCCTCCACAGTCTAGGATTACAAGCGTG 97917
RESULT 3
AC005098 AC005098 230552 bp DNA PRI 07-OCT-2000
LOCUS Homo sapiens BAC clone CTA-350L10 from 7q11.2, complete sequence.
DEFINITION AC005098
ACCESSION AC005098.2 GI:9211526
VERSION HTG.
KEYWORDS human.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 230552)
AUTHORS Sulston, J.E. and Waterston, R.
TITLE Toward a complete human genome sequence
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE 99063792
REFERENCE 2 (bases 1 to 230552)
AUTHORS Threide, J., Abbott, A., Graves, T., Elliott, G. and Markovic, C.
TITLE The sequence of Homo sapiens BAC clone CTA-350L10
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 230552)
AUTHORS Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (12-JUN-1998) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 4 (bases 1 to 230552)
AUTHORS Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (15-JUL-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 5 (bases 1 to 230552)
AUTHORS Waterston, R.
TITLE Direct Submission
JOURNAL Submitted (07-OCT-2000) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Jul 15, 2000 this sequence version replaced gi:3212893.
COMMENT ----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: sapiens@watson.wustl.edu
----- Summary Statistics

Center project name: H_RG350L10

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/GRN/CHR7>, send mailto:sgreen@nhgri.nih.gov, or see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

Clone CTA-350L10 is from a release of the human BAC library CITB-HS-A. The library contains cloned DNA from human sperm. See: Shizuya et al., Proc. Natl. Acad. Sci. USA 89:8794-7 (1992); U-J. Kim et al., Genomics 34:213-8 (1996). The clone is available from Research Genetics, Inc. (<http://www.resgen.com>).

VECTOR: pBelobAC11

Selection: chloramphenicol

NEIGHBORING SEQUENCE INFORMATION:

Actual start of this clone is at base position 1 of CTA-350L10; actual end is at base position 230552 of CTA-350L10.

The sequence CTA-350L10 from base position 222330 to 222775 is a GA rich region. The sequence is not exact but it is believed to be the best representation of this region. The region was sized by PCR from clone DNA at 650 bp. The region corresponds to restriction digest hindiii: band size 7685 in silico and 7756 real.

FEATURES

source	Location/Qualifiers
	1..230552
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="7"
	/map="7q11.2"
	/clone="CTA-350L10"
	/clone_lib="CITB-HS-A"
	441..725
repeat_region	/rpt_family="Alu"
	783..1149
misc_feature	/note="similar to EST BE000775 (NID:g8261008)"
misc_feature	910..1312
misc_feature	/note="similar to EST R22057 (NID:g776838) yh25f08.s1"
misc_feature	1286..1716
repeat_region	/note="similar to EST H72453 (NID:g1044269) ys04f07.r1"
	1609..1864
	/rpt_family="Alu"
misc_feature	1984..2271
misc_feature	/note="similar to EST R22113 (NID:g776894) yh25f08.r1"
misc_feature	2337..2682
misc_feature	/note="similar to EST AW383168 (NID:g6887827)"
misc_feature	2437..2558
repeat_region	/note="similar to EST N56845 (NID:g1200735) yw83e04.r1"
	2332..2601
repeat_region	/rpt_family="MIR"
	3054..3367
	/rpt_family="Alu"
misc_feature	3240..3391
misc_feature	/note="similar to EST A1566941 (NID:g4525393) tn25b07.x1"
misc_feature	3388..3845
misc_feature	/note="similar to EST H70120 (NID:g1040326) ys01a09.r1"
misc_feature	3599..4017
misc_feature	/note="similar to EST H89808 (NID:g1080238) yu82g03.r1"
misc_feature	3741..3928
misc_feature	/note="similar to EST N56845 (NID:g1200735) yw83e04.r1"
misc_feature	3816..3932
misc_feature	/note="similar to EST AW523936 (NID:g7166321)"
misc_feature	3819..3932
misc_feature	/note="similar to EST AW503416 (NID:g7139965)"
misc_feature	3820..3932
misc_feature	/note="similar to EST AW197058 (NID:g6476288) xb15f09.x1"
misc_feature	3820..3932
misc_feature	/note="similar to EST AW163398 (NID:g6302431) au94g04.y1"

Query Match

99.0%; Score 3267.6; DB 9; Length 230552;

Best Local Similarity 99.8%; Pred. No. 0;

misc_feature	3871..3932	/note="similar to EST A1700706 (NID:g4988606) we39e12.x1"
misc_feature	3876..3932	/note="similar to EST T31784 (NID:g613882)"
misc_feature	3889..4387	/note="similar to EST T79157 (NID:g697666) yd70b07.s1"
misc_feature	4109..4387	/note="similar to EST A1696334 (NID:g4984234) tw59e05.x1"
repeat_region	4502..4812	/rpt_family="Alu"
repeat_region	5041..5087	/rpt_family="Mariner"
misc_feature	5254..5330	/note="similar to EST AA794004 (NID:g2856959) vr37a02.r1"
misc_feature	5265..5316	/note="similar to EST AW503416 (NID:g7139965)"
misc_feature	5265..5330	/note="similar to EST A1700706 (NID:g4988606) we39e12.x1"
misc_feature	5265..5330	/note="similar to EST AW197058 (NID:g6476288) xb15f09.x1"
misc_feature	5265..5330	/note="similar to EST AW523936 (NID:g7166321)"
misc_feature	5265..5330	/note="similar to EST T31784 (NID:g613882)"
misc_feature	5285..5348	/note="similar to EST AW163398 (NID:g6302431) au94g04.y1"
misc_feature	5284..5330	/note="similar to EST T30644 (NID:g612742)"
misc_feature	5305..5330	/note="similar to EST AW993840 (NID:g8254025)"
repeat_region	5750..6041	/rpt_family="Alu"
misc_feature	5964..6127	/note="similar to EST AA078486 (NID:g1837960)"
misc_feature	6032..6370	/note="similar to EST A1566941 (NID:g4525393) tn25b07.x1"
repeat_region	6572..6878	/rpt_family="Alu"
misc_feature	6637..6883	/note="similar to EST AW748900 (NID:g7663832)"
misc_feature	6884..7119	/note="similar to EST AW895654 (NID:g8059859)"
misc_feature	6906..7117	/note="similar to EST AA601344 (NID:g2434969) no16a03.s1"
repeat_region	6939..6985	/rpt_family="MIR"
repeat_region	7127..7218	/rpt_family="L1"
repeat_region	7219..7425	/rpt_family="Alu"
repeat_region	7426..7476	/rpt_family="L1"
misc_feature	7510..7697	/note="similar to EST AA985239 (NID:g3163764) am80d09.s1"
misc_feature	7514..7697	/note="similar to EST T31784 (NID:g613882)"
misc_feature	7514..7697	/note="similar to EST AW523936 (NID:g7166321)"
misc_feature	7514..7689	/note="similar to EST A1700706 (NID:g4988606) we39e12.x1"
misc_feature	7514..7695	/note="similar to EST T30644 (NID:g612742)"
misc_feature	7514..7695	/note="similar to EST AW197058 (NID:g6476288) xb15f09.x1"
misc_feature	7514..7697	/note="similar to EST AW993840 (NID:g8254025)"
misc_feature	7514..7697	/note="similar to EST AA794004 (NID:g2856959) vr37a02.r1"
misc_feature	7565..7697	/note="similar to EST Z43001 (NID:g570579)"

	Matches	3293;	Conservative	0;	Mismatches	4;	Indels	3;	Gaps	2
Qy	1	tactaaaaatacaaaaattagccagcgctgggtggcgcacacctgtaatccccagctacttgg	60							
Db	47036	TACTAAAAATACAAAATTAGCCAGCGGTGGCGCACACCTGTAAATCCAGCTACTTTGG	47095							
Qy	61	gaagctgaggcagagagaatcgttgaacctggaaggcagaaggttgcagtgagccgagattc	120							
Db	47096	GAAGCTGAGGCAGAGAGATCGCTTTGAACCTTGGAGGCGAGAGGTTGTCAGTGAGCCGAGATT	47155							
Qy	121	gtgcactgcactccagcctgggcacaaagagcgaaacttcgcttcaacacaaataaatta	180							
Db	47156	GTGCACCTGCACCTCCAGCTTGGCAACAAGAGCGAAACTTCGCTTCACACAAATAAATTA	47215							
Qy	181	acgccacagatgtcttggctttcatctgccagaaactcaacctcaacccccaggagatcag	240							
Db	47216	ACGCCACAGCATGTCTTGGCTTTTCATCTGCCAGACCTCAACCTCACCCCCAGGAGATCAG	47275							
Qy	241	gtccggaccatgagctgacctgactcagcaagggtgagttggtgcagccctggcctcg	300							
Db	47276	GTCCGGACCATGAGCTGACCTTGACTCAGCGAAGGTTAGTTGGTGCAGCCCTGGCCCTG	47335							
Qy	301	ctgggaaggcagagctgcagcaggctgctgggctgagggcccgccactcatgaactcat	360							
Db	47336	CTGGGAGGCACAGGCTGCAGCAGGCTGCTGGGCTGAGGCCGCGCCACTCATGAACTCAT	47395							
Qy	361	gaccttaaatgagctcaaaaagctctggcctccaggctccaggctctaggggagtggaagaag	420							
Db	47396	GACCTTGAAATGAGCTCCAAAGCTCTGGGCTCCGAGGCTCTAGGGGAGTGGGAGAG	47455							
Qy	421	aggcctcagcctgtccctgggcagctgctgccccctctcactctttgtcccaaatccctc	480							
Db	47456	AGGCTCAGCGCTGCCCTGGGCACTGTGCCCCCTCCTCACCTCTTTGTCCCAATCCCTC	47515							
Qy	481	tcctggcaaaactgacagctttaatacaactctggagaaaaactgagtcagccctaaaggaa	540							
Db	47516	TCCTTGGCAAAAGCTGACAGCTTTAATATCACCTCTGGAGAAAAGCTGAGTCAGGCCCTAAAGGAA	47575							
Qy	541	caattcaatgaaccatttgccttacttgaggattggaaactcaagctctcactcaaatctgt	600							
Db	47576	CAATTCAATGAACCATTTGCTTACTTGGAGGATTGSAACTCAAGTCTCACTCAAAGTCTGT	47635							
Qy	601	gcaattttcgtccagctgtcactggccctcatccacacacaccaagaatgagcatcta	660							
Db	47636	GCCATTTTCGTCCAGCTGTCACTGGCCCTCATCCACACACACCAGGATCAGCATCTA	47695							
Qy	661	acgtctgatgcacactccactgcccggcttcattcactcattcattcattcattcactc	720							
Db	47696	ACGCTTGATGCACACTCCCATGCCGGTTCAATCACTCATTCATTTCATTCACTC	47755							
Qy	721	atcaattgactcaattcaattcaactcaactcaattcaactcaagtgaattgtgcagtc	780							
Db	47756	ATTCAATTGACTCAATTCAATTCACTCACTCACTCACTCACTCACTCACTCACTCACTC	47815							
Qy	781	acgtaccaaatattatggcctctgtgtgcagcacactagatgaagggctcgggctaga	840							
Db	47816	ACGATCCAAATATTATAGCCCTCTGTGTGCCAGCACATAGATGAGGGGCTGGGGCTAGA	47875							
Qy	841	gccctgataaccgggtcatgccctagctttcctctgggcacacacatttggtaaggggaga	900							
Db	47876	GCCCTGATAACCCGGTCATGCCCTAGCTTCTCTGGGACACACATTGTGTTAAGGGGAGA	47935							
Qy	901	ctaaaaaaataagtcaaggccagcgagtggtctcattgcctgaattccccagcactttggga	960							
Db	47936	CTAAAAAAATTAAGTCAGGCCAGCACGGTGGCTCATGCCCTGAATCCACAGCACTTTGGGA	47995							
Qy	961	ggccgagcgagtgaaattacctgaggtcagagttcgaagcacagcctggtgcacaaatgag	1020							
Db	47996	GGCCGAGCGGAGTGAATTAATCTGAGGTGACAGGTTCAAGACCGCCTGGCCCAACATGAG	48055							
Qy	1021	aaacccagctctctaataaaaaaataaaaaaaattagccaggtgtgggtggcacatgcctg	1080							
Db	48056	AAACCCAGCTCTTAATT--AAAAAATAAAAAAAATTAACCCAGGTGTGGTGGCACAATGCTGT	48113							

QY 2161 cggccaccccttgagtcacccagggccacacagctccagtgaggactaacgagggaggtct 2220
|||||
Db 49193 CGGCCACCCCTTGCACTGCCAGGGCCACACAGCTCCAGTGAGGACTAACGGGCGAGGCT 49252
QY 2221 tgggcacctggtccctggtcttgagcctgagactacccctgactccctgggaagacag 2280
|||||
Db 49253 TGGGCACCTGGTCCCTGGTCTTTGAGGCTGGATCTACCCCTCTGATCCCTGGGAAGACAG 49312
QY 2281 ttccttggaccgcccctggcccccagccctttactgtcccccgcctgtgtcccccagccag 2340
|||||
Db 49313 TTCCCTTGGACCCGCCCTGGGCCCCACGCCCTTTACTGTCTCCCGCCTGTGTCCCAGCCAG 49372
QY 2341 gccctagccttagcaggagctctcttctgtctccctgcccagtgccagcagccagc 2400
|||||
Db 49373 GCCCTCAGCCTTAGCCAGAGTCTCTTTCTGTCTCCCTGCCATGGCCAGCCAGCCAGC 49432
QY 2401 gctctccaggtccgagccacctctccaggaagcctccctgactagccagctatca 2460
|||||
Db 49433 GCTCTCTCAGGTCGGAGGCCACTCTCTCCAGGAAGCCTTCCCTGACTAGGCCAGCTATCA 49492
QY 2461 gagagtgcctcccaagagggagcctggaaactaaagctctctctctcccccagctgcc 2520
|||||
Db 49493 GAGAGTGGCCCTCCCAAGAGGAGGCTGGAACTAAGCTCTCTCTCCCGAGCTGCC 49552
QY 2521 ttagtgctcagttagtagctcttctccagtaggtgacacacatgacagggggccaata 2580
|||||
Db 49553 TGTAGTCTCAGTTAGAGTCTTATCTCTCCAGTAGGTGACACCATGACAGGGGCCAATA 49612
QY 2581 gagtctcccatctgtcccaagaggtcgagcaaaatgctgtctcagacacacaaagtcca 2640
|||||
Db 49613 GAGTCTCCCATCTGTGCCCAAGAGGCTGGCAAAATGCTCTGCTCAGACACACAAGTCCA 49672
QY 2641 ctgggtctccctaataccataggaagccagggaggaactacatttaggaatgaaagctt 2700
|||||
Db 49673 CTGGTCCCTCATCCATAGGAGGCCAGGAGAACTACATTTAGAAATGGAAGCTT 49732
QY 2701 gtatggaaacttagtctctgtgccaagacctctctcttttttttttttttttttttt 2760
|||||
Db 49733 GTATGGAACATTTAGTCTATGTGCCAAGACCTTCTCTTTTGTATTTTGTGTTT 49792
QY 2761 ttgagacagagctctgactgtgtccagccagagtgagtgagcagatctcagctcac 2820
|||||
Db 49793 TTGAGACAGAGTCTTGATCTGTGTGCCAGGCCAGAGTGCAGTGCAGATCTCAGCTCAC 49852
QY 2821 tgaacctccgcttccaggttcaactggttctctcctcagcctcagctcagagtagttggg 2880
|||||
Db 49853 TGCACCTCGCCTTCAGGTTCACTGGTTCTCCCTGCTCAGCTCCAGAGTAGTTGGG 49912
QY 2881 attacaggtgccaccaccacgcctggttaatttttttttttttttttttttttttttt 2940
|||||
Db 49913 ATTACAGGTGCCACACACCGCTGGCTAATTTTGTATTTTGTAGTAGAGAGGGTTT 49972
QY 2941 caccatgttgccagactggtctcaactcctgacctcagctcagtgatccaccacctgggcc 3000
|||||
Db 49973 CACCATGTTGGCCAGACTGTGCTCAAACTCCTGACCTCAAGTATCCACCACCTGGGCC 50032
QY 3001 tcccaagtgctgggattacaggcatgagccagcctgctgctgctgctgttttttttttttt 3060
|||||
Db 50033 TCCCAAGTCTGGGATTACAGCATGAGCCACCGCTGGCTGTGTTTGTGAAATGA 50092
QY 3061 ggtctggagtgcagtggtgcatcatagttcactgcagcctcaacctcccgagcccaagt 3120
|||||
Db 50093 GGTCTGGAGTGCAGTGTGGGATCATAGTTCACTGACGCTCGACCTCCAGGCCCAAGT 50152
QY 3121 gatctcctcctcagcccttagtagtgctggggtctacaggccacacacacatgctcgg 3180
|||||
Db 50153 GATCCTCTGCCCTCAGCCCTTGAGTAGCTGGGGCTACAGCGGCACACACCATGCTGTG 50212
QY 3181 ctagtttttaaaatttttggagatgaggtttcaactatgtttccagagctaattctgaa 3240
|||||
Db 50213 CTAGTTTTTAAAAATTTTGTGGAGATGAGGTTTCACTATGTTGTCCAGGCTAATCTTGA 50272
QY 3241 ctctcggcttaagcaacctctggtctcagcctccacagtgctaggtattacaagcgtg 3300

Db 50273 CTCCTCGCTTAAGCAACCTCTGTGCTCAGCTCCACAGTGTAGGATTACAAGGTG 50332
|||||
RESULT 4
AC004883 131359 bp DNA PRI 21-DEC-1999
LOCUS Homo sapiens PAC clone RP4-771P4 from 7q11.21-q11.23, complete
DEFINITION sequence.
AC004883
AC004883.2 GI:4263746
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 131359)
AUTHORS Sulston,J.E. and Waterston,R.
TITLE Toward a complete human genome sequence
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE 99063792
REFERENCE 2 (bases 1 to 131359)
AUTHORS Kalicki,J. and LaPlant,Y.
TITLE The sequence of Homo sapiens PAC clone RP4-771P4
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 131359)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (12-JUN-1998) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 4 (bases 1 to 131359)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (20-FEB-1999) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE 5 (bases 1 to 131359)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (18-MAR-1999) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE 6 (bases 1 to 131359)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (21-DEC-1999) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT On Feb 24, 1999 this sequence version replaced gi:3309099.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: sapiens@wustl.wustl.edu
----- Summary Statistics

Center project name: H_DU0771P04

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
All regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the

Db	93758	TTCCCTTGGACCCGCCCTTGGCCCCAGCCCTTTTACTGTCCCCGCCGTGTGTCCCCAGCCAG	93817
Qy	2341	gccctcagccttagccaggagtcctcttctgtctccctcgccatggccaggagcagccagc	2400
Db	93818	GCCCTCAGCCTTAGCCAGGAGTCCTTTCTGTCTCCCTGCCATGGCCAGCAGCCCCAGC	93877
Qy	2401	gctctctcaggtccgagggcccaetctctccaggagccttccctgactagccagctatca	2460
Db	93878	GCTCTCTCAGTCCGAGGCCCACTTCCTCCAGGAAGCCCTCCCTGACTAGCCAGCTATCA	93937
Qy	2461	gagagtggccctccaaagagggaagcctggaaaactaaagctctctctccccagctgcc	2520
Db	93938	GAGAGTGGCCCTCCCAAGAGGAGGCGCTGGAACATAAGCTCTCTCTCCGCCAGCTGCC	93997
Qy	2521	tgtagtgcagttagagtccttatcctctccagtagggtgcacacatgacaggggccaata	2580
Db	93998	TGTAGTGTCAGTTAGAGTCTTATCCTCTCCAGTAGGGTGACACCATGACAGGGGCCAATA	94057
Qy	2581	gagtcctcccatctgtctcccaaggaggctggacaaaatgcctgctcagacacacaagtcca	2640
Db	94058	GAGTCCCTCCCATCTGTCCCAAGGAGGCTGGACAAATGCCTGTCTCAGACACACAAGTCCA	94117
Qy	2641	ctgggtcccttaattcccataggaagccagggaagaaactacatttagaaaattgaagctt	2700
Db	94118	CTGGTCCCTTAATCCCATAGGAAGGCGAGGAGGAACATACATTTAGAAATTAAGAGCTT	94177
Qy	2701	gtaggaacatttagtccctatgtgccaaagacctctctcttttggttattttttgtgtt	2760
Db	94178	GTATGGAACATTTAGTCTATGTGCCAAGACCTTTCTCTTTTGTATTATTTTGTGTGT	94237
Qy	2761	ttgagacagagcttgatctgtgtgccaggccagagtgacgtggcagcagatctcagctcac	2820
Db	94238	TTGAGACAGAGTCTTGATCTCTGTGCCAGGCCAGAGTGCAGTGGCAGCATCTCAGCTCAC	94297
Qy	2821	tgcacctccgccttcagagttcaactggttctcctcctcagcctccaaagttagttggg	2880
Db	94298	TGCAACCTCCGCCCTTCAGAGTTCAACTGTGTCTCCTGCCTCAGGCTCCAGAGTAGTTGGG	94357
Qy	2881	attacagtgccccaccaccacgcctggctaattttggatttttagtagacagagggttt	2940
Db	94358	ATTACAGTGTGCCACCAACACACGCCTGGCTAATTTTGTATTTTAGTAGACAGGGTTT	94417
Qy	2941	cacatgttgccagagctggtctcaactcctgcactcaagtgcataccaccacactgggcc	3000
Db	94418	CACCATGTGGCCAGAGCTGGTCTCAAACTCCTGACCTCAAGTGATCCACCACCTGGGCC	94477
Qy	3001	tcccaaatgctgggattacaggcatgagccaccgtgcctggcctggttttttgaaatga	3060
Db	94478	TCCCAAGTGTGGGATTACAGGCATGAGCCACCGTGCCTGGCCCTGTTTTTTTGAATGA	94537
Qy	3061	ggctcgagtgagtggttgagatcatagtttcactgcagcctcaacctcccaggccccagt	3120
Db	94538	GGCTGGAGTGCAGTGGTGCGATCATAGTTCACTGACGCCCTCGACCTCCAGGCCCAAGT	94597
Qy	3121	gactcctgctcagcccccttgatagcttggggtacaggcgcacaccacacatgcctgg	3180
Db	94598	GATCCTCTGCTCAGCCCCCTTGAGTAGTGGGGCTACAGCGGCACACCATGCGCTGG	94657
Qy	3181	ctagtttttaaaattttgtgagatgaggtttcactcatgtgttccaggctaaactctgaa	3240
Db	94658	CTAGTTTAAAAATTTTGTGGAGATGAGGTTTTCATCTATGTTGTGCCAGGCTAATCTTGAA	94717
Qy	3241	ctctcggcttaagcaacctctggtctcagcctcccacagtgctcagattacaagcgtg	3300
Db	94718	CTCCTCGGCTTAAGCAACCCCTCTGGCTCAGCCCTCCCACTGCTAGATTTACAAGGGTG	94777

RESULT	5
AC004166	
LOCUS	AC004166 275197 bp DNA PRI
DEFINITION	Homo sapiens Chromosome 7 BAC Clone 239c10, complete sequence.
ACCESSION	AC004166
VERSION	AC004166.12 GI:8887011

KEYWORDS	HTG.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE	1 (bases 1 to 275197)
JOURNAL	Ren,Q., Burian,D., Huang,E., Meadows,S., Korenberg,J. and Roe,B.A.1
REFERENCE	Homo sapiens Chromosome 7 BAC Clone 239c10
AUTHORS	Unpublished
TITLE	2 (bases 1 to 275197)
JOURNAL	Burlan,D.M. and Roe,B.A.
REFERENCE	Direct Submission
AUTHORS	Submitted (21-FEB-1998) Department Of Chemistry And Biochemistry,
TITLE	The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
JOURNAL	OK 73019, USA
REFERENCE	3 (bases 1 to 275197)
AUTHORS	Burlan,D.M., Ren,Q., Meadows,S., Huang,E., Korenberg,J. and
TITLE	Roe,B.A.
JOURNAL	Direct Submission
REFERENCE	Submitted (01-JUL-2000) Department Of Chemistry And Biochemistry,
AUTHORS	The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
TITLE	OK 73019, USA
JOURNAL	4 (bases 1 to 275197)
REFERENCE	Ren,Q., Burian,D., Huang,E., Meadows,S., Korenberg,J. and Roe,B.A.1
AUTHORS	Direct Submission
TITLE	Submitted (20-JUL-2000) Department Of Chemistry And Biochemistry,
JOURNAL	The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
REFERENCE	OK 73019, USA
COMMENT	On Jul 1, 2000 this sequence version replaced gi:8779478.
FEATURES	Location/Qualifiers
source	1..275197
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="7"
	/clone="239c10"
BASE COUNT	70874 a 66648 c 66737 g 70938 t
ORIGIN	
Query Match	99.0%; Score 3265.6; DB 9; Length 275197;
Best Local Similarity	99.8%; Pred. No. 0;
Matches 3292; Conservative	0; Mismatches 4; Indels 4; Gaps 2
Qy 1	tactaaaatacaaaattagccaggcggtgtgagcgacacacctgaatcccgactacttgg 60
Db 33415	TACTAAAATACAAAATTAGCCAGGCGTGGTGGCGCACACCTGTAAATCCGACTACTTGG 33474
Qy 61	gaagctgaggcaggagaatcgcttgaaacctggaaggcagaggttcagctgagccgagatt 120
Db 33475	GAAGCTGAGGCAGGAGAAATCGCTTGAACCTGGAAGGCAGAGGTTGACGTGAGCCGAGATT 33534
Qy 121	gtgccactgcactccagctggcgcaacagagcgcgaacttcgcttcacaaacaaataaatt 180
Db 33535	GTGCCACTGCACATCCAGCTTGGGCAACAAGCGCAAACTTCGCTTCAACAAATAAAATTA 33594
Qy 181	acgccagcagtcttggtcttcatctgcagacctcaacctcaccgccagggatcag 240
Db 33595	ACGCCCAGCAGTCTTGGCTTTCATCTGCCAGACCTCAACCCCTCACCCCCAGGAGATCAG 33654
Qy 241	gtccggaccatgactgacctggactcaggcaagggtgagttggtgcagccctggcctg 300
Db 33655	GTCCGGACCATGAGCTGACCTGGACTCAGGCAAGGTTGAGTTGTCGAGCCCTGGCCTG 33714
Qy 301	ctggaggcagcaggtctgcagcagctgctggtggctgaagcccgcaactcatgaactcat 360
Db 33715	CTGGAGGCGCAGCGCTGCAGCAGGCTGCCCTGGGGCTTGAGGCCCGCCACTCATGAACTCAT 33774
Qy 361	gacctgaatgagctcctcaaaagctctggccctccagagctctaggggagtgaggagag 420
Db 33775	GACCTTGAAATGAGCTCCAAAGACTCTGGGCCCTCCACAGGCTCTAGGGGGAGTGGGAGAG 33834
Qy 421	aggcctcagcctgtccctgggcatgctgcccctcctcacctctttgtcccaaatccctc 480

Dbb 33835 AGGCTCAGCCTGTGCTTGGGCGATGCTGCCCCCTCCCTCACCTCTTTGTCCAAATCCCT 33894
QY 481 tcttgcaaaagctgacagctcttaataatacaactctgagaaactgagtcagccctaagaa 540
Dbb 33895 TCTTGGCAAAAGCTGACAGTCTTAATATCACTCTGGAGAAACTGAGTCAGCCCTAAGGAA 33954
QY 541 caattcaatgaaccatttcttacttgagattggaactcaactcaactcaactcaacttctgt 600
Dbb 33955 CAATTCAATGAACCAATTGCTTACTTGAGGATTGGAATCAAGTCTCACATCAAAAGTCTGT 34014
QY 601 gcaattttgtccagctgtcaactggccctcaatccacacacaccccaaggatgagcatcta 660
Dbb 34015 GCCATTTCTGCTCCAGCTGCTACTGGCCCTCATCCACACACACCAAGGATGAGCATCTA 34074
QY 661 acgcttgatgcacactcccatgcccggcttcattcaactcaactcaactcaactcaactc 720
Dbb 34075 ACGCTTCGATGCACATCCCATGCGCGGTTCATTCACTCATTCATTCAATTCATCACTC 34134
QY 721 attcaattgaactcattcaactcaactcaactcattcaactcaactcaactcaacttggcagtc 780
Dbb 34135 ATTCAATTGACTCATTCATTCACTCACTCACTCATTCATTCACTCAGTGAATGTTGCAGTC 34194
QY 781 acgatcaaatatttatggcctctgtgtgccaggcaactagatgagggctgggctaga 840
Dbb 34195 ACGATCCAAATATTTATGGCTCTGTGTGCCAGGCATAGATGGAGGGGCTGGGGCTAGA 34254
QY 841 gccctgataacccggtcatgccctagcttctctgggacacacattgtgtaaggagaga 900
Dbb 34255 GCCCTCATTAACCCGGTATGCCCCAGCTTTCTTGGGACACACATTGTGTTAAGGGGAGA 34314
QY 901 ctataaaaaattaaagtcaggcgagcggtggctcatgcttgaatccacagcaacttggga 960
Dbb 34315 CTAAAAAATTAAGTCAGGCAGGCAGCGTGGCTCATGCTGTAATCCAGCATTTGGGA 34374
QY 961 ggcgcagggcagtgaaattacactgaggtcagaggttcaagaccagcctggccaacatggag 1020
Dbb 34375 GGCCGAGCGCGAGTGAATTAAGTCTGAGGTGAGGAGTCAAGACCAAGCCCTGGCCAACTGGAG 34434
QY 1021 aaacccagctcttaataaaaaaataaaaaataagccaggtgtgtgtgcacatgctgt 1080
Dbb 34435 AAACCCAGTCTCTAATTT--AAAAAATAAATAATTAACAGGTGTGTTGGCACATGCTGT 34491
QY 1081 taatccagctactcaggagactaaccaagagaaattgcttgaaccagagagcgagaggt 1140
Dbb 34492 TAATCCAGCTACTCAGGAGACTAACCAAGAGAATTTGCTTGAACCCAGGAGGCAGAGGT 34551
QY 1141 tgcggtgagcagagatgcgcccattgcaactccagcctgggaacaaagagcgagactccat 1200
Dbb 34552 TGGCGTGAGCGGAGATCGGCCATTGCACTCCAGCTGGGAAACAAGAGCGAGACTCCAT 34611
QY 1201 ctcaaaaaaataaagtgagagcgagcgagagatcactagagccagtagtttga 1260
Dbb 34612 CTC-AAAAAATAAATAAGTGGAGCGCAGGAGGAGGATCACTAGAGCCAGTAGTTTGA 34670
QY 1261 gaccatctgggcaacatagcaggaccctgtctgtacaaaaataaaaaattaaataa 1320
Dbb 34671 GACCATCTGGCAACATAGCAGGACCTGTCTGTACAAAAAATAAATAAATAAATTAAC 34730
QY 1321 cgggca tgggtggcaacacccgtagtctccagctactcagaggtgagcgagagatcg 1380
Dbb 34731 CGGGCATGGTGGCACACACCCGTAGTCCAGCTACTCCAGAGGTGAGGAGGAGATCG 34790
QY 1381 ctggagccagaggttggaggtcagtgactgtgacttgcacccacactgacttaagcctg 1440
Dbb 34791 CTGAGGCCAGGAGTTGGAGGCTGCATGAACGTGTATCCACACACTGCGCTTAAGCCTG 34850
QY 1441 gatacaaaagcaagaccctgtctcaataaacaatagcaataataaataaataaataa 1500
Dbb 34851 GATAACAAAGCAAGACCCCTGTCTCAATAACATAGCAATAATAATAAATAAATAAATAA 34910
QY 1501 tgcatttgcagatgcacagtgataagctgtctgcagaaaaagagcgaggaagagctg 1560
Dbb 34911 TGCAATTTGCGATGCATCAGTGATAAGTGCTTCGACAAAAAGGAGGAGGAGGAGGCTG 34970

QY 1561 aaaaagtatgaggtttgctatgcaatgtgaattatcaaggaaggcttctcgaaagag 1620
Dbb 34971 AGAAAGGTATGAGGTTTGCTATGCAATGTGAAGTTATCAAGGAAGGCTTCTCGGAAGAGG 35030
QY 1621 tgacatttgagcagagaataatggagagagttatggaggaagatggtgaattggggggaac 1680
Dbb 35031 TGACATTTGAGCAGAGAAATGGAGGAGATTATGGAGGAAGATGGTCAATGGGGGAAC 35090
QY 1681 atgctaaagaccagaataatgtgtcaaggggggaaagatggtcaaggggggacagcaaatg 1740
Dbb 35091 ATGTCAAGACACAGGAATATGTTCAAGGGGGAAAGATGGTCAAGGGGACACGCAAAATG 35150
QY 1741 caaaggccctgagcgagagcagctgattcaaccccaaaacccgtggggccgtgcag 1800
Dbb 35151 CAAGGCCCTGAGGCGAGGAGGAGCTTGAATCACCCCCCAAAACCCGTGGGGCCGTGCAGG 35210
QY 1801 cgagggaaggaacaagtgttaaaccttctctgtccctgcaggtgtgtgtgaacatgag 1860
Dbb 35211 CGACGGGAAGGACAAGTGAACCCCTTTTCCCTGCGAGGTGTGTGTGAACATGAG 35270
QY 1861 tctcccatgtttacacccctgaagcctgaagctgaaagctcccccaaaactgaaagaagaacaa 1920
Dbb 35271 TCTCCCATGTTTACACCCCTGCAAGCCTGAAGAGTCCCCAGAAACTGAAGAAGAAGCA 35330
QY 1921 agccttctgtacccctctgccccctgtccgacgcgcacaaaagcagcttcccttt 1980
Dbb 35331 AGCCCTTCTGTACCTTCCCTGCCCCCTGTCCGACCCGACAAAAGCGACTTCTCTCTTT 35390
QY 1981 ccagtgtattaaaggcgagcctggaagtgcaggagagcactgaggccaccagtcag 2040
Dbb 35391 CCAGTGCATTTAAGCGCGAGCCTGGAAAGTGCAGGGAGCACTGGAGGCCACCCAGTCATG 35450
QY 2041 ggggaacacctcaatccgtcaacatgcacctgtgggtttggagagcgcttcgtaccacagc 2100
Dbb 35451 GGGGACACCTTTCATCCGTCACATCGCCCTGTGTGGGCTTTGAGAAGCGCTTGTGATCCACAGC 35510
QY 2101 cagcacatgtgagtagctgtgagggcatccgtgggggaaatcacgggagggagacaga 2160
Dbb 35511 CAGCACTATGTGAGTAGCTGTGTGGGGCATCCCGTGGGGGGAATACGGGAGGGACACGA 35570
QY 2161 cgggcaaccttgagtcaccaggggcaaccagctccagtgaaggactaacggggcgaggtct 2220
Dbb 35571 CGGCAACCTTGCAGTCCAGGGGCAACAGCTCCAGTGAGGACTTAACGGGCGCAGGGTCT 35630
QY 2221 tgggcaacctgtccctggttcttgagcctggatctacccctctgtatccctggggaagacag 2280
Dbb 35631 TGGGCACTTGGTCTGCTGTGTTGAGCCTGGATCTACCCCTCTGATCCTCGGAAGACAG 35690
QY 2281 tcccttggaccccgcttggggcccgagccttactgtccccgcctgtgtccccagccag 2340
Dbb 35691 TTCCCTTGGACCCCGCTTGGGCCCCAGGCCCTTACTGTCCCCCGCTGTGTCCCCAGCCAG 35750
QY 2341 gccctcagccttagccaggagctctcttctgtccccctgcaatggccagcgagccccagc 2400
Dbb 35751 GCCCTCAGCCTTAGCCAGGAGTCTCTTTCTGCTCCCCCTGCCATGGCCAGCAGCCACAGC 35810
QY 2401 gctctctcaggttcgagggccactctctccaggaagccttccctgactagccagccta tca 2460
Dbb 35811 GCTCTCTCAGGTCCGAGGCCCACTCTCCAGGAAGCCTTCCCTGACTAGCCAGCATCA 35870
QY 2461 gaggatggccctcccaagagagggagcctgaaataaagctctctctctccccagctgac 2520
Dbb 35871 GAGAGTGGCCCTCCCAAGAGGGAGGCCCTGGAAACTTAAAGCTCTCTCTCTCCCCAGCTGCC 35930
QY 2521 tgaagtgcagttagagcttctatctctccagtaggggtgacacacatgacaggggccaata 2580
Dbb 35931 TGTAGTGTCACTAGAGTCTTATCTCTCCAGTAGGGGTGACACCATGACAGGGGCCAATA 35990
QY 2581 gagtctctccatctgtcccccaagaggtgtgacaaaatgcctgtctcagacacacagtcaca 2640
Dbb 35991 GAGTCTCCCATCTGTCCCCAAGAGGGGTGGCAAAATGCCCTGCTCAGACACACAAGTCCA 36050

Db	1021	CAGGGCCAAACAGCTCCAGTGAAGACTAAAGGGCAGGGTCTTGGGACACTGGTCCCTGG	1080
Qy	2239	tctttgagcctggatctacccctctgatacccttgggaagacagtctccttggaccgcctc	2298
Db	1081	TCCTTTGAGCCTGGATCTACCCCTCTGATCCCTGGGAAGACAGTTCCCTGGACCCGCCCT	1140
Qy	2299	ggccccagcccttactgttcccggcctgtgtccccagccagggccctcagccttagccag	2358
Db	1141	GGGGCCAGGCCCTTTACTGTCCCGCTGTGTCCAGCCAGGCGCCCTCAGCCTTAGCCAG	1200
Qy	2359	gagtcctcttctgtccctgcattggccaggcagcccaagcgtctctcaggtccgagg	2418
Db	1201	GAGTCTCTTTTCTGCTCCCTGGCATGGCCAGGACGCCAGCGCTCTCTCAGGTCCGAG	1260
Qy	2419	cccactctccaggaagccttccctgactagccagactatcagagagtgggccctcccaag	2478
Db	1261	CCCACCTCTCCAGGAAGCCTTCCCTGACTAGCCACGCTATCAGAGAGTGGCCCTCCCAAG	1320
Qy	2479	agggaggcctggaaactaaagctctctctccccagctgcctgtgtagtgtagttagtgc	2538
Db	1321	AGGGAGGCTTGGAAACTAAAGCTCTCTCTCTCCACAGCTGCCATAGAGTCTCTCCCATCTGTCC	1380
Qy	2539	ctatactctccagttaggtgacacaaatgacaggggccaaatagagtctctccaatctgtcc	2598
Db	1381	CTTATCTCTCTCCAGTAGGGTGACACCATGACAGGGGCCAATAGAGTCTCTCCCATCTGTCC	1440
Qy	2599	ccaagaggctggacaaatgccctgctcagacacacaagtccactgggtccccctaaatccca	2658
Db	1441	CCAAGAGGCTGGACAATGCCCTGCTCAGACACACAAGTCCACTGGGTCCCTTAATCCCA	1500
Qy	2659	taggaagggcaggaggaaactacattaggaaattgaagcttgatggaaacatttagtcc	2718
Db	1501	TAGGAAGGCCAGGAGGAACATACATTTAGGAAATTGAAGCTTGTATGTGAACATTTAGGCC	1560
Qy	2719	tatgtgcaagaccttctctctttttttgttatattttttgtttttgttttgagacagagcttgat	2778
Db	1561	TATGTGCCAGACACTTCTCTTTTGTATTATTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	1620
Qy	2779	ctgtgtcccaggccagagtgagtgagcagatctcagctcaactgcaacctccgccttcca	2838
Db	1621	CTGTTGCCAGGCCAGAGTGGAGTGGACCATCTCAGCTCAGTGCACAACTCCGCCCTTCCA	1680
Qy	2839	ggttcaactggttctctcgtccctcagctccagagtagttggggattacaggtgccccacc	2898
Db	1681	GGTTCAACTGGTCTCTCGCTCAGCCTCCAGAGTAGTTGGGATTTACAGGTACCCACCAC	1740
Qy	2899	cacgcctggctaattttgttattttttagtagagacaggggtttcaccaatgttggcagact	2958
Db	1741	CACGCCCTGGCTAATTTTGTATTATTTTATAGTAGACACAGGGTTTCCACCATGTTGGCCAGACT	1800
Qy	2959	ggctcaaaactccttgacctcaagtgtatccaccacctgggcctcccaaaagtgcctggagtt	3018
Db	1801	GGTCTCAAACTCTTGACCTCAAGTAGTATCCACCACCTTGGCCCTCCCAAGTGTGGGATTT	1860
Qy	3019	acaggcatgagccaccgtcgtcgtggcgtgttttttggaaatgaggtctggagtgagtcgtt	3078
Db	1861	ACAGGCATGAGCCACCCTGCTGGCCCTGTTTTTTTGAATGAGGTCTGGAGTGAAGTGGT	1920
Qy	3079	gcgatactagttaactgcagcctcaaacctcccaggcccaagtgtactctcgtcctcagcc	3138
Db	1921	CGATATATAGTTACTTCAGGCCCTCAAGCTCCAGGCGCCAAAGTAGTATCTCTGTGCTCAGCC	1980
Qy	3139	ccttgagttagtggggtcagagcgacacacacatgcctggcgttagtcttttaaaattttt	3198
Db	1981	CCTTGAGTAGTGGGGCTCAGGTGGACACACCATGCTGGCTGGCTAGTTTTTAAAAATTTTT	2040
Qy	3199	gtggagatgaggttttactatgttgtccaggtcaaatcttgaactcctcgtgcttaagcaac	3258
Db	2041	GTGGAGATGAGTTTCACTATGTTGTCCAGGCTTAATCTTGAACTCCTCGGTTTAAGCAAC	2100
Qy	3259	cctctgtctcagcctcccacagtgctgaggtattacaaggtg	3300

Db 2101 CCTGTGCTCAGCCTCCACAGTGTAGGATACAGCGTG 2142

RESULT 9
HS47PHOX01
LOCUS
DEFINITION Homo sapiens p47-phox (NCF1) gene, exons 1-5.
ACCESSION U57833
VERSION U57833.1 GI:2754709
KEYWORDS
SEGMENT
SOURCE 1 of 3
ORGANISM human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 8131)
AUTHORS Gorlach,A., Lee,P.-L., Roesler,J., Hopkins,P.J., Christensen,B.,
Green,E.D., Chanock,S.J. and Curnutte,J.T.
TITLE A p47-phox pseudogene carries the most common mutation causing
p47-phox- deficient chronic granulomatous disease
J. Clin. Invest. 100 (8), 1907-1918 (1997)
97474758
MEDLINE
REFERENCE 2 (bases 1 to 8131)
AUTHORS Chanock,S.J., Roesler,J., Hopkins,P., Lee,P., Bassett,D.T.,
Christensen,B., Curnutte,J.T. and Gorlach,A.
TITLE Genomic Structure and the Identification of Multiple Polymorphisms
in the p47-phox Gene
Unpublished
REFERENCE 3 (bases 1 to 8131)
AUTHORS Gorlach,A., Roesler,J., Christensen,B., Chanock,S.J. and
Curnutte,J.T.
TITLE Direct Submission
JOURNAL Submitted (09-MAY-1996) P. Lee, Molecular and Experimental
Medicine, Scripps Research Institute, 10550 North Torrey Pines Rd,
La Jolla, CA 92037, USA
FEATURES
source
1. .8131
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="7"
/clone="p1 clone 42"
/map="7q11.23"
877. .948
/gene="NCF1"
/number=1
4143. .4223
/gene="NCF1"
/notes="this is defined as the wildtype gene since it
contains GTGT at the beginning of exon 2"
/number=2
5957. .6032
/gene="NCF1"
/number=3
6132. .6297
/gene="NCF1"
/number=4
7657. .7712
/gene="NCF1"
/number=5
BASE COUNT 1835 a 2214 c 2088 g 1994 t
ORIGIN

Query Match 53.8%; Score 2105.8; DB 9; Length 8131;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 2132; Conservative 0; Mismatches 7; Indels 2; Gaps 2;

Qy 1161 ccattgacctccagcctgggaacaagacgagcagctccatctcaaaaaaaagtg 1220
|||||
Db 1 CCATTGCACTCCAGCCTGGGAACACAGAGCGAGACTCCATCTC-AAAAAAGTGG 59
|||||

Qy 1221 gaggcagagcagagagatcactagagccagtagtttgaccatcctgggaacatag 1280
|||||

Db 60 GAGCAGAGCAGCAGGATCACTAGAGCCAGTAGTTTGAGACCATCTGGGCAACATAG 119
Qy 1281 caggacctgtctgtacaaaaaattaaaaaatttaacccgggcatgtggcacacacc 1340
|||||
Db 120 CAGGACCTGTGTGTACAAAAAATTTAAACCGGCATGGTGGCACACACC 179
Qy 1341 cgtlgtccccagctactccagagcgtgagcaggaggtcgtgagcccgaggttggag 1400
|||||
Db 180 CTTAGTCCCAGCTACTCCAGAGGCTGAGGAGGATCGCTGGAGCCAGGAGTTGAG 239
Qy 1401 gctgcagtgaactgtgatccccaccactgcacttaagcctggataacaagaagaccctg 1460
|||||
Db 240 GCTGCAGTGAATGTGTATCCACCATGCGCTTAAGCCTGGATAAACAAGCAGACCCCTG 299
Qy 1461 tctcaataacaatagcaataataaagaataataaataaataaattgaattgcagtccatcag 1520
|||||
Db 300 TCTCAATAACAATAGCAATAATAATAAAGAAAAATTTAAATGCAATTTGCGATGCATCAG 359
Qy 1521 tgataagctctgcagaaaaagcagcaggaagcgtgagaaaggtatgaggtttgct 1580
|||||
Db 360 TGATAAGTCTCTGCAGAAAAAGGAGCGCAGAGAGGCTGAGAAAGGTATGAGGTTTGTCT 419
Qy 1581 atgcaatgtgaagtattcaaggagggtctctcggaagaggtgacatttgagcagagaaat 1640
|||||
Db 420 ATGCAATGTGAAGTTATCAAGGAAGGCTTCTCGAAGAGGTTGACATTTGAGCAGAGAAAT 479
Qy 1641 ggaggagatttgagggaagatgtgaatggggggaacatggtcaagaccaggaatat 1700
|||||
Db 480 GGAGGAGAGTTATGGAGGGAAGATGGTGAATGGGGGAACATGGTCAAGACCAGGAATAT 539
Qy 1701 ggtcaaggggggaaaagatggtcaaggggagcagcaaatcaaaagccctgagcaggag 1760
Db 540 GTTCAAGGGGGGAAAGATGTTCAAGGGGAGCGCAGCAAAATGCAAAAGCCCTTGAGGCAGGAG 599
Qy 1761 cagcttgatcaccccccaaacccgtgggcccgtgcaggcagcgggaagacaagtgt 1820
Db 600 CAGCTTGATTACACCCCAAAACCCGTGGGCCCGTCGAGCGACGGGAAGACAAAGTGA 659
Qy 1821 aaccttttctgtccctgcagggtgtgtgaacatgagctctgccatgtttacacct 1880
Db 660 AACCTTTTCTGTCTCCTGCAGGTGTGTGAACATGAGTCTGCCCATGTTTACACCTT 719
Qy 1881 gcaagcctgaagagtcctcccgaaaactgaagaagaagaacaaagcccttctgtacctccc 1940
Db 720 GCAGCCTGAAGAGTCCCCAGAACTGAAGAAGACGAAAGCCCTTCTGTACCCCTCCC 779
Qy 1941 tgcacctgtcccgaccgcgacaaaagcgaacttctcttccagtcatttaagcgcag 2000
Db 780 TGCCCTCTGTCCCGACCGCGACAAAGCGACTTCTCTCTTCCAGTGCATTTAAGGCGCAG 839
Qy 2001 cctggaaagtccagggagcactggagccaccagtcattggggacaccttcatccgtca 2060
Db 840 CCTGGAAGTGCAGGAGGACACTGGAGGCCACCCAGTATGCGGGGACACCTTTCATCCGTCA 899
Qy 2061 catcgccctgctgggctttgagaagcgttcgtaccacgagcagcactatgtgagtagctg 2120
Db 900 CATCGCCCTGCTGGGCTTGTGAGAGCGCTTCGTACTCCAGCAGCACATATGTGAGTAGCTG 959
Qy 2121 gtggagggcat -cccgtggggggaatacagggaggttcgtcccccagccaccttgagtcacc 2179
Db 960 GTGGAGGGCATCCCGCTGGGGGAATACGGGAGGAGACGACGCCACCTTTCAGTGTCC 1019
Qy 2180 agggccaacagctccagtgagaactaacgggcaggtcttgggaccttggtccctcagt 2239
Db 1020 AGGCCCAACAGCTCCAGTGAGGACTAACGGGCGAGGGCTTTGGGGACCTTGGTCCCTTGGT 1079
Qy 2240 ctttgagccttggaactacccctctgacctcctgggaagacagattcccttggaccgccctg 2299
Db 1080 CTTTGGACCTGGATCTACCCCTCTGTATCCCTTGGGAAGACAGTTCCTTGGACCCGCCCTG 1139
Qy 2300 ggcacccagcccttactgtccccgcctgtgtccccagccagccctcagccttagccagg 2359
Db 1140 GGCCCCAGGCTTTTACTGTCCCCCTGTGTCCCCAGCCAGGCGCTCAGGCTTAGGCCAGG 1199

exon

1. >1965
/gene="NCF1"
/pseudo
/note="contains 5' UTR"
/number=1

BASE COUNT	551 a	503 c	517 g	394 t
ORIGIN				
Query Match	58.9%	Score 1943.4;	DB 9;	Length 1965;
Best Local Similarity	99.6%;	Pred. No. 0;		
Matches 1958;	Conservative 0;	Mismatches 6;	Indels 1;	Gaps 1;
Qy	74	gagaatcgttgaacctggaagcgagaggttgcagtgagccgagattgtgccactgcact	133	
Db	1	GAGAACTCCTTTGAACCTTGAAGCGAGAGAGGTTGCACTGAGCGCGAGATTGTGCCACTTGCAC	60	
Qy	134	ccagctggcgaaacagcgaaacttcgcttcaaacataataataacgcccagcagctg	193	
Db	61	CCAGCTGGGCAACAGAGCGAATCTTCCTTCAACAAATAATAATTAACGCCACAGATGT	120	
Qy	194	cttggtcttcattctgcagacactcaacccctcacccccaggagatcaggtcccgaccatga	253	
Db	121	CTTGGCTTTTCATCTGCCAGACTCAACCCCTCACCCCCAGGAGATCAGTCCGGACCACGA	180	
Qy	254	gctgacctgactcaggaagcgaggttgagttggtgcagccctggcctgctggaggacag	313	
Db	181	GCTGACCCCTGGACTCAGCAAGAGGTGAGTTGGTGAGCGCCCTGGCTGTGGGAGGCACAG	240	
Qy	314	gctgcagcagctgcctgggctgagggccgcacctcatgaactcatgaactcatgaactgaatga	373	
Db	241	GCTGCAGCAGAGCTGCCTGGGGCTGAGGCCACACACTCATGAACCTCATGACCTTGAATGAG	300	
Qy	374	ctccaaaagctctgggctccaggtctagggggagtgggagagagaggcctcagcctg	433	
Db	301	CTCCAAAAGCTCTGGGGCTCCAGGCTCTAGGGGAGTGGGAGAGAGGCTCAGGCCTG	360	
Qy	434	tccttggcatgctgccccctcctcaacctttgtccaaatcccccttccttggcaaaagt	493	
Db	361	TCCCTGGGCATGCTGCCGCCCTCCACACTCTTTGTCCAAATCCCTTCTCTGGCAAGCT	420	
Qy	494	gacagctcttaatacaactctggagaaaactgagtcagccctaagaagacaaatcaatgaac	553	
Db	421	GACAGCTCTAATATCACTCTCGAGAAACTGAGTCAGGCCCTAAGGAACAATTCATGTAAC	480	
Qy	554	catttgcttacttgaggatggaaactcaagtctcaactcaaaagtctgtgccatttctgccc	613	
Db	481	CATTTGTCTTACTTTAGGATTTGAACTCAAGTCTCACTCAAAAGTCTGTGCCATTTTCGTCC	540	
Qy	614	cagctgctactggccctcaatcacacacacccaaggaatgagcatctaaagccttgcatgca	673	
Db	541	CAGCTGTCACTTGGGCCCTCATCCACACACACCAAGGATGAGCATCTAAGGTTGCATGCA	600	
Qy	674	cactcccatgcccggttcattcactcattcattcattcattcactcattcattcattcattc	733	
Db	601	CACCTCCATGCCCCGGTTCAATTCATCTATTCAATTCATTCATCTATTTCATTCAC	660	
Qy	734	ttcattcattcactcactcattcattcactcactcactcactcactcactcactcactc	793	
Db	661	TTCAATTCATTTCACCTCACTCATTTCAATTCATTCAGTGAATTTGCAAGTCAAGAT	720	
Qy	794	ttatggcctctgtgtccagcactatagagggtggggcttagagccctgataacc	853	
Db	721	TTATGGCCCTCTGTGTGCCAGGCATAGATGAGGGGCTGGGGCTTAGAGCCCTCTGATAACC	780	
Qy	854	cggtcatgcccctagctttcctgggacacacattgtgtgtaaggggagacataaaaaattaa	913	
Db	781	CGGTCACTGCCCTAGCTTTCCCTGGGACACACATTTGGTTAAGGGAGACTAAAAAATTA	840	
Qy	914	gtcaggccagcgacgggtggctcattgctgaaatccacagcaacttttggaggccgaggtc	973	
Db	841	GTCAGGCCAGCAGGTTGGCTCATTCCTTGAATCCAGACATTTGGAGGCCGAGGCCAGT	900	

Qy	974	gaattac	ctgagtgctcagggattcaagaccagcctggtgccaacatggaagaaacccagttctct	103
Db	901	GAATTACCT	GAGTCAAGGAGTTCAAGACCAGCCTGCCAACATGGAGAAACCCAGTCTCT	960
Qy	1034	aatttaaaaaa	aaaaaaaaaattagccagggtgtgtgtggcacatgcctgttaatcccaagctac	1093
Db	961	AATTAAAAA	AAAAAAAAAAAAAATTACCCAGGTGTGTGTGGCACATGCCCTGTAAATCCCAAGTAC	1020
Qy	1094	tcagagact	aaacgaagagaattcttgaaccagagagcagaggttcggttgagccga	1153
Db	1021	TcAGGAGACT	TACGCNAAGAGAAATGCTTGAACCCAGAGAGCAGAGTTCGGTGTGAGCCGA	1080
Qy	1154	gatcgcgcca	tgtcactccagcctggtgaaacaagagcgagactccactctc-aaaaaaaa	1212
Db	1081	GATCGCGCA	TTCGACTCCAGCCTGGGAAACAAGAGCGAGACTCCATCTCAAAAAA	1140
Qy	1213	aaaagtggag	gcagagggagaggtactaagagccagtagtttgagaccatcctgtgg	1272
Db	1141	AAAAGTGG	AGGCGACAGAGGAGGAGTACACTAGAGGCCAGTACTGTTTGAGACCATCTCTGGG	1200
Qy	1273	caacatagc	agagccctgctctgtacaaaaaaattaaaaaaatttaaccggcgcatggtg	1332
Db	1201	CAACATAGC	AGGACCCCTGCTGTGTACAAAAAATTAATAAAAAATTTAACCGGGCATGTGG	1260
Qy	1333	cacacccgt	tagtcccgactctccagaggtcgagcagaggtatcgctggagcccg	1392
Db	1261	CACACCCG	TAGTCCCAGCTACTCCAGAGCTGAGCAGGAGATCGCTGGAGCCAGG	1320
Qy	1393	agttggagg	ctcgagtgaactgtgatccaccactgcacttaagcctggatacaaaagca	1452
Db	1321	AGTTGGAG	CTCGAGTAACAAATAGCAATTAATAAAGAAAAATTAATGCAATTTGCGA	1380
Qy	1453	aqacccctg	ctcaaatataacatagcaataataataaagaaaaaattaaatgcatttgcga	1512
Db	1381	AGACCCGT	GTCTCAATAAACAATAGCAATTAATAAAGAAAAATTAATGCAATTTGCGA	1440
Qy	1513	tgcatcagt	tgaataagtgctctgcagaaaaagggcaggaagaggtcgaaaggtatga	1572
Db	1441	TGCATCAG	TGATAGTGCTCTGCAGAAAAAGGAGGCAGGAAGAGGTGAGAAAGGTATGA	1500
Qy	1573	ggttgtcta	tgcaatgtgaagttaatacagaagggcttctcggaagagtgacatttbgac	1632
Db	1501	GGTTTTGT	ATGCAATTGAGGTTATCAAGGAAGGCCTTCGGAAGAGGTGCATTTGAGC	1560
Qy	1633	aqagaaatg	aagagagttatagagggaagatgtgaatggggggaacaatggtcgaagacc	1692
Db	1561	AGAGAAAT	GGAGAGAGTTATGGAGGGAAATGGTGAATGGGGGGAACATGGTCAAGACC	1620
Qy	1693	aggaata	tggtcaaggggggaagaatggtcgaagggagcagcaaatgcaaaagccctga	1752
Db	1621	AGGAATAT	TGGTCAAGGGGGGAAGATGCTCAAGGGAGCGCAGCAATATGCAAAAGCCCTGA	1680
Qy	1753	ggcagagc	agcttgattcaccccccaaacccgtggggccgtgagcagggacaggggaagga	1812
Db	1681	GGCAGGAG	CAGCTTGATTACCCCCAAAAACCCGTGGGGCCGTGCAGGCCACGGGAAGGA	1740
Qy	1813	caagtgtaa	acccttttctctgtccctgcaggtgtgtgtgaacaatgagctgcccattgtt	1872
Db	1741	CAAGTGTAA	ACCCCTTTCTCTGTCCCTGCAGGTGTGTGAACATGAGTGTGCCCATGTT	1800
Qy	1873	tacacccct	gaagctggaaggtccccagaaactgaagaaagcaaaagcccttctgt	1932
Db	1801	TACACCTG	CAGCCTGAAGAGTCCCCAGAACTGAAAGAGAGCAAGCCCTTTCTGT	1860
Qy	1933	acccttctc	tggccctgtccccgacgcgcacaaagcacttctcttctccagtgcat	1992
Db	1861	ACCCCTCG	CCCTGTGCCCGTGTCCCGACCGCGACAAAGCGACTTCTCTTTCCAGTGCATTTA	1920
Qy	1993	aggcgagc	ctggaagtgcagggagcactgagggccacccagtc	2037
Db	1921	AGGCGAGC	CTGGAGTGCAGGGAGCACTGGAGGCCACCCAGTC	1965

Db 1381 TTATCTCTCCAGTAGGTCACACCATGACAGGGGCCAATAGAGTCTCTCCCATCTGTGCC 1440

QY 2600 caaggagctggacaaatgctgtcagacacacacaaagtcactgggtccctcaatcccat 2659

Db 1441 CAAGGAGGCTGGACAAATGCTGTCTCAGACACACAAAGTCCACTGGTCCCTTAATCCCAT 1500

QY 2660 aggaagccagggaggaactacatttaggaaattgaagctgtgatgaaacatttagtctc 2719

Db 1501 AGGAAGCCAGGGAGGAACATACATTTAGGAATTTGAAGCTGTATGGAACATTTAGCCCT 1560

QY 2720 atgtgcca 2728

Db 1561 ATGTGCCAA 1569

RESULT 13

HS47P43S01

LOCUS HS47P43S01 7258 bp DNA PRI 07-JAN-1998

DEFINITION Homo sapiens p47-phox (NCF1) pseudogene, clone P43, exons 1-5.

ACCESSION U60970

VERSION U60970.1 GI:2754714

KEYWORDS 1 of 3

SEGMENT human.

SOURCE Homo sapiens

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 7258)

AUTHORS Gorlach,A., Lee,P.L., Roesler,J., Hopkins,P.J., Christensen,B., Green,E.D., Chanock,S.J. and Curnutte,J.T.

TITLE A p47-phox pseudogene carries the most common mutation causing p47-phox- deficient chronic granulomatous disease

JOURNAL J. Clin. Invest. 100 (8), 1907-1918 (1997)

MEDLINE 97474758

REFERENCE 2 (bases 1 to 7258)

AUTHORS Chanock,S.J., Roesler,J., Hopkins,P.J., Lee,P.L., Bassett,D.T., Christensen,B., Curnutte,J.T. and Gorlach,A.

TITLE Genomic structure and the identification of multiple polymorphisms in the p47-phox gene

JOURNAL Unpublished

AUTHORS 3 (bases 1 to 7258)

TITLE Gorlach,A., Lee,P.L., Roesler,J., Chanock,S.J. and Curnutte,J.T. Direct Submission

JOURNAL Submitted (17-JUN-1996) Molecular and Experimental Medicine, The Scripps Research Institute, 10550 N. Torrey Pines Rd., CAL 1, La Jolla, CA 92037, USA

FEATURES

source Location/Qualifiers

1..7258

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="P1 clone P43"

/chromosome="7"

/map="7q11.23"

/note="Human genomic P1 clone P43 is a p47-phox pseudogene which is defined by the presence of a GT deletion at the beginning of exon 2"

<19..90

/gene="NCF1"

/pseudo

/number=1

3252..3330

/gene="NCF1"

/pseudo

/number=2

5084..5159

/gene="NCF1"

/pseudo

/number=3

5259..5424

/gene="NCF1"

/pseudo

/number=4

6784..6839

/gene="NCF1"

/pseudo

/number=5

BASE COUNT 1574 a 2009 c 1839 g 1836 t

ORIGIN

Query Match 38.3%; Score 1263.6; DB 9; Length 7258;

Best Local Similarity 99.6%; Pred. No. 0;

Matches 1277; Conservative 0; Mismatches 4; Indels 1; Gaps 1;

QY 2020 actgaggccaccagtcacatggggagacaccttcacgtccatcgccatcgccctctgggtttt 2079

Db 1 ACTGGAGGCCACCCAGTCATGGGGGACACCTTCATCCGTCACATCGCCCTCTGGGCTTT 60

QY 2080 gagaagcgtcttgaccagccagcactatgtgagtagctgtaggggcat-ccgtgg 2138

Db 61 GAGAAGCGCTTCGTACCCAGCCAGCAGCTATGTGAGTAGCTGTGGAGGGCATCCCGGTGG 120

QY 2139 ggggaatacagggagggagcagcagggccaccttgcagtcaccaggggccaaccagctccagt 2198

Db 121 GGGGAATACGGGAGGGAGCAGCAGCGCCACCTTGCAGTCCCGAGGCCAACCCAGCTCCAGT 180

QY 2199 gaggactaaagggcagggtcttgggcaactggtccctgggtcttggagcctggatctacc 2258

Db 181 GAGGACTAACGGGCGAGGGTCTTGGGCACCTGGTCCCTGGTCTTTGAGCCTGGATCTACC 240

QY 2259 cctctgacctgggaagacagattcccttggaaccgcctgggcccagccctttactgt 2318

Db 241 CCTCTGATCCCTGGGAAGACAGTTCCCTTGGACCGCGCTGGGCCCGCCAGCCCTTTACTGT 300

QY 2319 cccgcctgtgtcccccagccagggccctcagcttagccagagtcctctttctgctcccc 2378

Db 301 CCCGCCCTGTGCCCGAGGCCCTCAGCCTTAGCCAGGAGTCTCTTTCTGCTCCCC 360

QY 2379 tgccatggccaggcagcccgctctctcaggtccaggtccagggcccactctccaggaagct 2438

Db 361 TGCCATGGCCAGGCGGCCAGCGCTCTCTCAGTCCGAGGCCACTCTCCAGGAAGGCT 420

QY 2439 tccctgactagccagctatcagagatggccctcccaagaggagccctgaactaaa 2498

Db 421 TCCCTGACTAGCCAGCTATCAGAGAGTGGCCCTCCCAAGAGGAGGCTTGGAACTAAA 480

QY 2499 gctctctctccccagctgctgtagtgtagttagtcttctctccctcagtaggt 2558

Db 481 GCTCTCTCTCCCCAGCTGCTGTAGTGTAGTGTAGTCTTATCCTCTCCAGTAGGCT 540

QY 2559 gacaccatgacaggggccaatagagtcctcccatctgtccccaaggaggtggacaaatg 2618

Db 541 GACACCATGACAGGGGGCCAAATAGAGTCTCTCCCATGTGCCCAAGGAGGCTGGACAAATG 600

QY 2619 cctgctcagacacacaagtccactgggtccctaatccataggaagccagggaggaac 2678

Db 601 CCTCTCAGACACACAGTCCACTGGGTCCCTTAATCCATAGGAAGCCAGGAGGAAC 660

QY 2679 tacatttagaaattgaagctgtgatgaaacatttagtcctatgtgccaaagcctttctc 2738

Db 661 TACATTTAGGAATTTGAAGCTTGTATGGAACATTTAGCCCTATGTGCCAAGACCTTTCTC 720

QY 2739 ttttttgttatttttgtgttttgagacagattgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2798

Db 721 TTTTGTGTATTTTGT 780

QY 2799 cagtggcacatctcagctcactgcaacctccgcttccaggttcaactgttctctcgc 2858

Db 781 CAGTGGCAGCATCTCAGCTCACTGCAACCTCCGCTTCCAGGTTCAGTGTCTCTCTGC 840

QY 2859 ctacgctccagagtagtgggattacaggtgtccaccaccacgcctgtgctaaattttgt 2918

Db 841 CTCAGCCTCCACAGTAGTTGGATTACAGGTGCCACCACCACGCCCTGGCTTAATTTTGT 900

QY 2919 atttttagtagagacaggggtttcaccatgttggcagactggtctcaaacctcctgacctc 2978

Qy 2799 cagtggcagcatctcagctcactgcaacccctccgcctccagggttcaactggtctctcgc 2858
|||||
Db 781 CAGTGGCAGCATCTCAGCTCAGTGCACACCTCCGCTTCAGAGGTCACTGGTCTCCTGC 840
Qy 2859 ctcagcctccagagtagtggtgattacaggtgcccacacacagcctggtaattttgt 2918
|||||
Db 841 CTCAGCCTCCAGAGTAGTGGAGCTACAGGTGCCACACACAGCCTGGCTAATTTTGT 900
Qy 2919 atttttagtagacagaggtttacacatggttgccagactgggtctcaactcctgacctc 2978
|||||
Db 901 ATTTTGTAGTAGACAGGGTTTACCATGTTGGCCAGACTGGTCTCAAACTCCTGACCTC 960
Qy 2979 aagtgtatccaccacactggcctcccaaaagtgtgggtattacaggtatgagccacgtgc 3038
|||||
Db 961 AAGTGATCCACCACCTGGCCCTCCCAAGTGTGGGATTTACAGGCATGAGCCACCTGTC 1020
Qy 3039 ctggcctgttttttgaatgaggtctggagtcagtgagtggtgctgcatcatagttcaactgcag 3098
|||||
Db 1021 CTGGCCTGTTTTTTGAAATGAGGTCTGGAGTGCAGTGGTGGGATCATAGTTCACTGCAG 1080
Qy 3099 cctcaactccagggcccaagtgatcctcctgctcagcccttgagtagctgggggtac 3158
|||||
Db 1081 CCTCAACTCCCAAGGCCAAGTATCCTCTGCTCAGCCCTTGAGTAGCTGGGGCTAC 1140
Qy 3159 aggcgcacacacacacactcctggctagtttttaaaattttgtggagatgaggtttcacta 3218
|||||
Db 1141 AGGTGCACACCACTGCTGGCTAGTTTTTAAATTTTTGTGGAGATGAGGTTTCACTA 1200
Qy 3219 tgtgtccaggttaatttgaactcctcgtgcttaagcaacccctctgctcagcctccca 3278
|||||
Db 1201 TGTGTCCAGGCTAATCTGAACCTCCCTGGTGTATGCAACCCCTCTGCTCTCAGCCTCCCA 1260
Qy 3279 cagtgtagattacaagctg 3300
|||||
Db 1261 CAGTGTAGGATTACAAGCGTG 1282

RESULT 15
HSNCF1S1
LOCUS HSNCF1S1 3594 bp DNA PRI 07-JAN-1998
DEFINITION Homo sapiens p47-phox (NCF1) pseudogene, clone P38, exons 1 and 2.
ACCESSION U69639
VERSION U69639.1 GI:2754730
KEYWORDS
SEGMENT 1 of 5
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 3594)
AUTHORS Gorlach,A., Lee,P.L., Roesler,J., Hopkins,P.J., Christensen,B.,
Green,E.D., Chanock,S.J. and Curnutte,J.T.
TITLE A p47-phox pseudogene carries the most common mutation causing
p47-phox- deficient chronic granulomatous disease
J. Clin. Invest. 100 (8), 1907-1918 (1997)
97474758
REFERENCE 2 (bases 1 to 3594)
AUTHORS Chanock,S.J., Roesler,J., Hopkins,J., Lee,P.L., Bassett,D.T.,
Christensen,B., Curnutte,J.T. and Gorlach,A.
TITLE Characterization of the genomic structure of the p47-phox gene
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 3594)
AUTHORS Gorlach,A., Lee,P.L., Roesler,J., Christensen,B., Chanock,S.J. and
Curnutte,J.T.
TITLE Direct Submission
JOURNAL Submitted (06-SEP-1996) Molecular and Experimental Medicine, The
Scripps Research Institute, 10550 N. Torrey Pines Rd., CAL 1, La
Jolla, CA 92037, USA
FEATURES
source Location/Qualifiers
1..3594
/organism="Homo sapiens"
/db_xref="taxon:9606"

/chromosome="7"
/map="7q11.23"
/clone="P38"
/note="pseudogene is defined by the presence of a GT
deletion at the beginning of exon 2"
<1..90
/gene="NCF1"
/number=1
/pseudo
3240..3318
/gene="NCF1"
/number=2
/pseudo
BASE COUNT 738 a 1030 c 843 g 983 t
ORIGIN

Query Match 37.9%; Score 1251.6; DB 9; Length 3594;
Best Local Similarity 99.6%; Pred. No. 0;
Matches 1265; Conservative 0; Mismatches 4; Indels 1; Gaps 1;

Qy 2032 ccagtcattgggggacacatttcacgtacatcgccctgctggcctttgagaagcgcttc 2091
|||||
Db 1 CCAGTCATGGGGGACACCTTCATCCGTCACATCGCCCTGCTGGCGTTTGAAGAAGCGCTTC 60
Qy 2092 gtacccagccagcactatgtagtagctggtgagggcat-cccgtgggggaaatcacagg 2150
|||||
Db 61 GTACCCAGCCAGCACTATGTGAGTAGCTGGTGGAGGGCATCCCCGTGGGGGAAATACCGG 120
Qy 2151 agggacagcagcggccaccccttgccagtcgccaggggcacaacagctccagtgaggactaacgg 2210
|||||
Db 121 AGGCACAGCAGCGGCACCCCTTGACAGTCCAGGGCCCAACAGCTCCAGTGAGGACTTAACGG 180
Qy 2211 ggcagggtttgggcaacctggctccctggtcttttggcctgagctgatatccctctgacct 2270
|||||
Db 181 GGCAGGGTCTTTGGCACCTGGTCCCTGGTCTTTGAGCCTGGATCTACCCCTCTGATCCCT 240
Qy 2271 ggggaagacagtctcccttgaccgccctggcccccagcccttactgtccccgcctggt 2330
|||||
Db 241 GGGGAAGACAGTTCCTTGGACCCGCCCTGGGCCCCAGGCCCTTTACTGTCCCGCGCTGTGT 300
Qy 2331 cccagccagccctcagccttagccaggagtcctcttctgtccctccctgcccagggccag 2390
|||||
Db 301 CCCAGCCAGGCCCTCAGCCCTTAGCCAGGAGTCTCTTCTGCTCCCTGCCATGGCCAG 360
Qy 2391 gcagccagcgtctctcaggtccagagggccactcctccaggaagccttcctgactagc 2450
|||||
Db 361 GCAGCCACAGCGCTCTCTCAGTCCGAGGCCCACTCCTCCAGGAAGCCTTCCCTGACTAGC 420
Qy 2451 ccagctatcagagagtggtgcccctcccaagagagggcctggaaactaaagctctctctc 2510
|||||
Db 421 CCAGCTATCAGAGAGTGGCCCTCCCAAGAGGGAGGCCCTGGAAACTAAAGCTCTCTCTC 480
Qy 2511 cccagctgctctgtagtgcagtttagagtcttctcctctccagtaggggtgacaccatgaca 2570
|||||
Db 481 CCCAGCTGCCCTGTAGTGTGACGTCTTATCCTCTCCAGTAGGGGTGACACCATGACA 540
Qy 2571 ggggccaatagagtcctcccatctgtccccaagagaggtggacaaaatcctgctcagaca 2630
|||||
Db 541 GGGGCCCAATAGAGTCCCTCCCATCTGTCCCCAAGGAGGCTGGACAAATGCCCTGCTCAGACA 600
Qy 2631 cacaagtcacagtgggtcccttaattcccatagaaagggccagggaggaactacatttagaa 2690
|||||
Db 601 CACAAGTCCACTGGGTGCCCTTAATCCCATAGAGGCCAGGGAGGAAGTACATTAGGAA 660
Qy 2691 attgaagctgtatggaaacatttagtccctatgtgccagaccttctctctttttgttatt 2750
|||||
Db 661 ATTCAGCTTCTATGGAACATTTAGCCCTATGTGCCAAGACCTTTCTCTCTTTTGTATT 720
Qy 2751 tttttgttttgagacagagcttgcattgttccccagggccagagtgagtgagcagcat 2810
|||||
Db 721 TTTTGTGTGTTTGGACAGAGCTTTGATCTGTGCCCCAGGCCAGAGTGCAGTGGCAGCAT 780

Qy	2811	ctcagctcactgcacaactcgcgcttcctccagggttcaactgggtttctcctgcctcagctccag	2870
Db	781	CTCAGCTCACTGCAACTCCGCCTTCCAGGTTCAACTGGTTCTCCTGCCTCAGCCTCCAG	840
Qy	2871	agtagttgggattacaggtgccccacacacagcctggctaattttttatttttagtaga	2930
Db	841	AGTAGTTGGGATTACAGGTGCCACCCACACGCCTGGCTAAATTTTGTATTTTAGTAGA	900
Qy	2931	gcacaggtttcaccatgttgccagactggctcaaaactcctgaactcctcaagtaatccacc	2990
Db	901	GACAGGGTTTCAACATGTGGCCAGACTGGTCTCAAACTCTGACCTCAAGTGATCCACC	960
Qy	2991	cacctgggctcccaagtgctgggattacagggatgagccacgctgctctggcctgttt	3050
Db	961	CACCTGGGCTCCCAAAGTGCTGGGATTACAGGCATGAGCCACCGTGCCTGGCCTGTTTT	1020
Qy	3051	tttgaatgaggtctbgagtgcgagtgggtgcgatcatagttcactgcagcctcaactccc	3110
Db	1021	TTTGAATGAGGTCTGGAGTGAAGTGGTGGCATCATAGTTCACTGCAGCCTCAACCTCCC	1080
Qy	3111	aggcccaagtatcctcctgcctcgaagcccttgagtgcagctgggctcacaggcgcaacca	3170
Db	1081	AGCCCCAAGTGATCTCTCCCTGCCTCAGCCCTTGAGTACGTGGGGCTACAGGTGCACACA	1140
Qy	3171	ccatgcctggctagttttcaaattttctggagataggtttcacatattgtctcaggc	3230
Db	1141	CCATGCCTGGCTAGTTTTTAAATAATTTTGTGGAGATGAGGTTTCACATGTTGTCAGGC	1200
Qy	3231	taacttggaactcctcctggcttaagcaacccctctggtctcagcctcccaagtcgtagat	3290
Db	1201	TAACTTGAACCTCTGGTTTAAAGCAACCCCTGGTCTCAGCCTCCACAGTGCTAGGAT	1260
Qy	3291	tacaagcgtg 3300	
Db	1261	TACAAGCGTG 1270	

Search completed: January 11, 2002, 16:31:36
Job time: 26648 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 11, 2002, 16:23:10 ; Search time 814.02 Seconds
(without alignments)
3539.803 Million cell updates/sec

Title: US-09-820-005-3_COPY_10200_13560
Perfect score: 3361
Sequence: 1 aaaggtcaggcactgtgc.....cttgccaagtgggtgcatt 3361

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_1101.*

- 1: /SID88/gcgdata/geneseq/geneseq/NA1980.DAT.*
- 2: /SID88/gcgdata/geneseq/geneseq/NA1981.DAT.*
- 3: /SID88/gcgdata/geneseq/geneseq/NA1982.DAT.*
- 4: /SID88/gcgdata/geneseq/geneseq/NA1983.DAT.*
- 5: /SID88/gcgdata/geneseq/geneseq/NA1984.DAT.*
- 6: /SID88/gcgdata/geneseq/geneseq/NA1985.DAT.*
- 7: /SID88/gcgdata/geneseq/geneseq/NA1986.DAT.*
- 8: /SID88/gcgdata/geneseq/geneseq/NA1987.DAT.*
- 9: /SID88/gcgdata/geneseq/geneseq/NA1988.DAT.*
- 10: /SID88/gcgdata/geneseq/geneseq/NA1989.DAT.*
- 11: /SID88/gcgdata/geneseq/geneseq/NA1990.DAT.*
- 12: /SID88/gcgdata/geneseq/geneseq/NA1991.DAT.*
- 13: /SID88/gcgdata/geneseq/geneseq/NA1992.DAT.*
- 14: /SID88/gcgdata/geneseq/geneseq/NA1993.DAT.*
- 15: /SID88/gcgdata/geneseq/geneseq/NA1994.DAT.*
- 16: /SID88/gcgdata/geneseq/geneseq/NA1995.DAT.*
- 17: /SID88/gcgdata/geneseq/geneseq/NA1996.DAT.*
- 18: /SID88/gcgdata/geneseq/geneseq/NA1997.DAT.*
- 19: /SID88/gcgdata/geneseq/geneseq/NA1998.DAT.*
- 20: /SID88/gcgdata/geneseq/geneseq/NA1999.DAT.*
- 21: /SID88/gcgdata/geneseq/geneseq/NA2000.DAT.*
- 22: /SID88/gcgdata/geneseq/geneseq/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	406	12.1	52216	22	AAH28355
C 2	391.4	11.6	32152	22	AAI57791
C 3	385.8	11.5	51474	20	AAF97846
C 4	382.8	11.4	119950	22	AAFX90201
5	375.2	11.2	1624	22	AAH98793
6	373.2	11.1	72604	20	AAZ10752
C 7	370.6	11.0	32367	19	AAV35620
C 8	370	11.0	1298	22	AAF64178
C 9	370	11.0	162450	21	AAZ86967
C 10	369	11.0	15577	19	AAV35616
C 11	367.6	10.9	32192	22	AAI63523
C 12	366.6	10.9	106746	21	AAA10225
C 13	366	10.9	32134	22	AAI63522
C 14	363.6	10.8	160552	22	AAD02697
C 15	363	10.8	236303	22	AAI11614
C 16	362.6	10.8	8427	22	AAI57755
C 17	359.6	10.7	160552	22	AAD02697
C 18	359.2	10.7	23603	22	AAI62936
C 19	358.4	10.7	4352	22	AAI58667
C 20	358.4	10.7	4377	22	AAI60453
C 21	358.4	10.7	72604	20	AAZ10752
C 22	357.4	10.6	31853	22	AAI63343
C 23	355.2	10.6	1887	22	AAI61210
C 24	354.8	10.6	5301	22	AAF97866
C 25	354.2	10.5	24218	22	AAI62935
C 26	353	10.5	110000	22	AAF84800
C 27	352.2	10.5	122186	22	AAC89560
C 28	351.6	10.5	4124	22	AAI1154
C 29	351.2	10.4	122186	22	AAC89560
C 30	350.4	10.4	16891	20	AAI37084
C 31	350.4	10.4	161425	22	AAH02340
C 32	350.4	10.4	162025	22	AAH02339
C 33	349.8	10.4	65656	21	AAI53450
C 34	349	10.4	23613	22	AAI62930
C 35	348	10.4	32204	22	AAI57790
C 36	347.4	10.3	3884	22	AAH98627
C 37	346.2	10.3	986	21	AAH79767
C 38	345.6	10.3	52216	22	AAH28355
C 39	345	10.3	66685	22	AAI07380
C 40	344.8	10.3	2562	22	AAH17646
C 41	344.8	10.3	114793	22	AAD08215
C 42	344.6	10.3	32249	22	AAI62932
C 43	343.6	10.2	2483	22	AAH16966
C 44	343.4	10.2	14919	22	AAI63992
C 45	343.2	10.2	32190	22	AAI62927

ALIGNMENTS

RESULT 1
AAH28355/C
ID AAH28355 standard; DNA; 52216 BP.
XX
AC AAH28355;
XX
XX 05-SEP-2001 (first entry)
XX
DE Nucleotide sequence of the human musashi promoter.
XX
XX Musashi promoter; multipotential neural progenitor cell;
KW neural stem cell; central nervous system; ss.
XX
XX Homo sapiens.
XX
XX WO200146384-A2.
XX
PD 28-JUN-2001.
XX
XX 22-DEC-2000; 2000WO-US35395.
XX
XX 23-DEC-1999; 99US-0173003.
XX
XX (CORR) CORNELL RES FOUND INC.
XX (NISC-) JAPAN SCI & TECHNOLOGY CORP.
XX
XX Goldman SA, Okano H;
XX
XX WPI; 2001-4i8053/44.
XX
XX Separating multipotential neural progenitor cells from a mixed
PT population of cells, involves introducing nucleic acid molecule
PT encoding fluorescent protein under promoter control, and separating
PT fluorescent cells -

XX Claim 12; Fig 11A-JJ; 87pp; English.

XX The present sequence represents the human musashi promoter, which is used in the method of the invention. The specification describes a method for separating multipotential neural progenitor cells from a mixed population of cell types. The method comprises introducing a nucleic acid encoding a fluorescent protein under control of a promoter which selectively functions in the progenitor cells, into all cell types of the mixed population, allowing only the progenitor cells to express the fluorescent protein, and identifying and separating cells that are fluorescent, where the separated cells are progenitor cells. The method is useful for isolation and purification of multipotential neural progenitor cells, especially neural stem cells from adult brain. The isolated cells are used in both basic analyses of precursor and stem cell growth control, as well as in more applied studies of their transplantability and engraftment characteristics. The cells are useful in support of the structural repair of the damaged central nervous system, such as in the traumatized brain, or the contoured, traumatized or transected spinal cord.

XX Sequence 52216 BP; 13494 A; 12461 C; 12362 G; 13899 T; 0 other;

Query Match 12.1%; Score 406; DB 22; Length 52216;
Best Local Similarity 54.3%; Pred. No. 3,4e-83;
Matches 1401; Conservative 0; Mismatches 1015; Indels 166; Gaps 22;

Qy 6 gtcaggcactgtggtcctgtaataccagcactttggaggccgagcggttgat 65
Db 11688 GCCGGCAGCGTGTCTCATGCTGTAATCCAGCACTTTGGGAGCGCGGTGGAT 11629

Qy 66 cacttgagtcaggatctgagaccagcctgacaaatctgtaactcc-atctctact 124
Db 11628 CAC--GAGTCAAGAGATCGAGACCATCTGGCTAACATGGTGAACCCCTGCTCTACT 11571

Qy 125 aaaaatacaaaaatagccaggtgtggtggcggtgtgcttagtccagctacttggag 184
Db 11570 AAAAATAAAAAATAGCTGGGCATGGTGTGTGCTATAGTCCAGCTACTTGGGAG 11511

Qy 185 gctgaggcaggtgaattacttgaacctggaaggtgaggtgtgcaatgagccaagattgca 244
Db 11510 GCTGAGGCAAGAGAATGCTTGAACCCGGAGGAAGAGGTTGTGTGAGCCCAAGATCATG 11451

Qy 245 ccactgcactcca-----gtgacagagcagactccactcaaaaaaaataaaaaa 298
Db 11450 CCAGTGCCTCCAGCCTGGGTGACAGAGCAAGACTTTGTCTCAAAAAAATAAAAAA 11391

Qy 299 agttgggaaagccaggtgagtgagtggtccagcgtgtaatcccaacactttaagagct 358
Db 11390 A---AAAAAAGGTCGTGGTGGCGTGGCT--GACGCCGTAATCCAGCACTTTGAGAGGCC 11335

Qy 359 gaggtaggaatcctttgagccagaggttcgagaccagcctgggcaattgtcccaagac 418
Db 11334 AAGGTGGGTGGATCA--TGAGATCAGGAGTTCAGACCAAGCCTACCCCAATATAGTGAAC 11277

Qy 419 ctgtcttttacaataaa-----ttagccgggtgtgtggcatacgtctgtgtccca 470
Db 11276 TCCGCTCTACTAAAAACACAAACTTAGTTGGGCACTGGTGGCGTGTGCGCTGTATGCCA 11217

Qy 471 gctattcggaggtgagcaggaggagattgtttgagccttaggagctgagggctgagtga 530
Db 11216 GCTGCTCCAGAGGCTGAGGTAAAGATAACACTTGAACCCAGGAGGAGGTTGCAGTGA 11157

Qy 531 gctgtgatacagctactgtactctagcctgagcagcagcagcagcagcagcagcagcagc 590
Db 11156 GCTAAGACTGCACTGCACTCCAGCTGGGCAACAGGCAAACTGTGCTCAAAAAA 11097

Qy 591 agaaaaataaagtgtgggaaggtcactcatcagatgagacaagacatgttga 650
Db 11096 AAAATTGTAATAATATAAATGTAATAATATATATATATAGAGAAACTAGAGGAAATAA 11037

Qy 651 agttagcgcgaagcctggagaacgctatgcgccaggaataatgcagggcagcagagact 710

Db 11036 ATTAAATGTTAATA-----TGGTTGCCTCAATAATATGGGGATTATATTT 10989

Qy 711 caagatgcagcgcgcgtgtcttgaggccagatggccctgcaatgcccaactcaacctgc 770
Db 10988 TATTTAATCAACACATATTTCTTTAAGAAACTGTGTTTATTTAATGACATTTTGTGTTT 10929

Qy 771 cctcccttggccacacatcacgcgcccccatcctcgacagctaccgcgcgcattgc 830
Db 10928 TTTGGTTTTTTTGGAGATGGAGTCTTGCTC-----TGTGCAAACTGGAGTGCAGTGGC 10873

Qy 831 cgactacgagaagcactcggtcctcgagatggctctgcccagggggagcgtggtgaggt 890
Db 10872 GAGATCTGCTCACTGCAACTCGCCCTCCAAAGTTCAAGCGATTTCTCTGCTCAGCC 10813

Qy 891 cgtgagaagcagcagcgtgcagacctcccaactccacacttccggggctccctccctgggtct 950
Db 10812 TCCTGAGTAG-CTGGGACTACAGGTGCGCCTCACCATGCCAGCTAAATTTTATATTTT 10754

Qy 951 caggaaaccacagccacaagccccctgccaaggctcaggcagcctggccccctgggagac 1010
Db 10753 AGTAGACAGGG-----TTTTACCATGTTGGATACAGGTG 10717

Qy 1011 tccagctgttagggccctaaatgtcctccccacactgtgggtcgctctctctcttag 1070
Db 10716 TGAGCCTCCGTGCCCGGCTGTATTTCTTTAAATAATATATATAGAATATATATAGC 10657

Qy 1071 tgtgcacctgtgtgtggtgtggtcgtgtgcagcgcggggcggggagcagcgtctgt 1130
Db 10656 AAAACACATTTCTTTGTTGTTTGTGTTGAGAGAGAGTCTTGTCTGTCAACCCAG 10597

Qy 1131 cgtgtctgtcgtggtggtatgggacgcgtctgtctcattatgaagtggcctcagagctgt 1190
Db 10596 GCTGATAG-----CAGTGGCATGATCTCTGCTCAGTGC 10563

Qy 1191 gattctgtgagcatgtgtgcatgcatgtgacatcattgtcagctgtgtggtgaggtg 1250
Db 10562 AACCTCTCCCTCCGGTTCAACAATTTCTCTGCTCAGCCTCCCAAGCAGCTGGGATG 10503

Qy 1251 acatttccaaatctgagcatgtgacatcagtgctgtctgtctgtccctgtgtcctcaccatcc 1310
Db 10502 ACAGCGCGCTGCCACCCCGCGGTAATTTTGTATTTTAGTAGAGATAGAGTTTTCAC 10443

Qy 1311 ctgattggtcagggagcgcgtggccctgcctccagtcacattcccgcaacctgtgca 1370
Db 10442 CATGTTGGCCAGGCTGTCTGAACTCTGACCTTAGCTGATCCACCCACCTCAGCCTCC 10383

Qy 1371 caggtgtgtgtctgtcagatgaaagcaagcagcagcgtggtgataccacacatccttctctg 1430
Db 10382 CAATGTGTAGG-----GATTACAGCATGAGCCTGATCCAGCCCTTAACACCAT 10331

Qy 1431 agccccctgagcagctcctgacgagacggaagaccctgagcccaactatcaggtgtccccct 1490
Db 10330 TTTCTATTAGTAGTTTTTTA-----AAAAATTACAAATGTAAGAT 10293

Qy 1491 gcccctccgaggtgtgaggggtgtgggagaaaggcagcagcagcagcagcagcagcagcagc 1550
Db 10292 GTGCTCACA-----TTAAAAAAGAAAAATCGCTCAGTACACAAAGGTATAAGATG 10243

Qy 1551 gactcgtttgagctgtgggt 1610
Db 10242 AAAAAAGTAAAGTTCTTACCCCTTAGACTCTCTCCCTTAAATGTAACTGTACCAATTTTC 10183

Qy 1611 tgggtctgt 1668
Db 10182 ATGTATATCTTCCAAAAAATGTTCCATGCTTATATCTTATAGATATATATATATTTT 10123

Qy 1669 agggagggagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1728
Db 10122 AAAAAATAGATATAGCTACTTGGGAGGCTGAGTGGGAGGATCACTTGAGCCCCAAGAGG- 10062

Qy 1729 caggtctcagtggaagcctgaaggatgagcagacottagcctcaggaagaggaggtgcctg 1788

RESULT	2	
AAI57791/c		
ID	AAI57791	standard; DNA; 32152 BP.
XX		
XX		
AC	AAI57791;	
XX		
DT	19-OCT-2001	(first entry)
XX		
XX		
DE	Human colorectal cancer antigen coding sequence	SEQ ID NO: 328.
XX		
XX	Human; colorectal cancer; colorectal cancer antigen; gene therapy; ds.	
KW		
XX		
XX		
OS	Homo sapiens.	
XX		
XX		

Qy	588	ggaatgcyagggcagcagagactcaagatgccagcgc-----ctgtctctgagggccca	740
Db	29504	CACGAATCCCTTACTGTGTAGACATTTTGTATTTGGCACCTTATCAGCTTTATACATGAGC	29445
Qy	741	qatggggccclgcaatgccactcaccctgcctccctctctgccccagacatcacccggccc	800
Db	29444	GATGCTCCCAAGTTGAGTTTAAAGAGAAATTCAGACCTTGCAGAGAGCCTCTGTAAAGAGCC	29385
Qy	801	catcatcctgcagcgtaccgcgcctatgcccactacgaaagacacctcgggctccgaga-	859
Db	29384	TAGGGAGGCCAGACGCCGTGGCTCACGCCCTGTAACTCTAACACTTTTGGGAGGCCGAGAC	29325
Qy	860	----tggctctgtccacgggggagcgtggtgaggtctgtgagaagagcgcgagcgtcag	915
Db	29324	AGGCAGATTGTCTGAGCTCAGGAGTTTGTAGACCAGCCAGCGCAACACGGGAAACCCCGT	29265
Qy	916	acctcccaaccttacgggggtcccttcccctgctcgaagaaacccacagccacaagccccc	975
Db	29264	CTCTACTTAAATAACAAAAAGTAGCCGGGGCTGGTGTGGGCACCTGTAGTCCCAGGTAC	29205
Qy	976	tgcgaagctcagcagccctggcccctggaggactccagctctgttaggggcccctaaat	1035
Db	29204	TCGGGAGGCTCAGCAGAGAAATGCTTGAACCCGGAGCGGAGGTTCAGTGCATCCAA	29145
Qy	1036	gtctcccccaactgtgggtgcctctctcttagtgtcacccctgtggtgctgtgggc	1095
Db	29144	GATGGCAC--CACTGTCACTCCAGCCTGGGTGCACAGAGCGAGACTCCGTCTCCAAAAAANA	29087
Qy	1096	atctgtgcattggcagccggggcggggcattgtctcgtgttctgtctcgtatgggtatggg	1155
Db	29086	AAAAAAGAGCCCAAGCGCACACTGCAAGAAGCCACATGCAGGGGTGAGGAGAGCTAGGC	29027
Qy	1156	accgtctgttctattgaagtgggctcagagctgtgattctgtgag-----ca	1203
Db	29026	CAGCTCAGGTCATCCTTAGGNAANAACCTTACCTGTCCATGGAACAGGTGACTGAGTGCCA	28967
Qy	1204	tgtgtgcattgcattgtgactcattgttccagtggtgtgaaggtgacatttcccaatc	1263
Db	28966	TCAGAGTTAGGGAGAGGTGGGCTGAGGAGAAATTTCTTCAGTTTGTCTCCAGGTGCAT	28907
Qy	1264	tgaacattggacatcagttgtctgtgttccctgtgtctcaccatccctgatggctgcag	1323
Db	28906	CCCCCTGTAGACTTTAGCCCCCCACAGAGAGATGGGTAGCATATAAAGTGAAGAAGCAAA	28847
Qy	1324	ggagccctggccctcccctcagtcacattccgcacctctgagcaggttggttggtt	1383
Db	28846	GTCCGGGATTTGGCTGGTGCCTGTAGTCTCAGCTGCTTGGGAGGTTGGGTGGGAGGAT	28787
Qy	1384	c---tgtcagatgaagcaagcgaggtggatccccagcatcctctcagcccccttga	1440
Db	28786	GGCTTTAGCCAGCAGATTGTGTCTGCCCTGGGCAACACAGCTTTATTAGGAAAAAGGA	28727
Qy	1441	cagtccttgacgcagcagaagacccttgagcccaactatgcaggtgccccctgccctccag	1500
Db	28726	AGGGGAAGGGGAAAGGAAGGAGAGAGAGGAAGGAAGGAGGAAGACAGACAGAGAGAGG	28667
Qy	1501	gctgtaggggtgtgggagaaaggggcgaggcagg-----ctca	1538
Db	28666	AGACAGTCAGANAAGAAAGGAAGGAGGAGGAAGGAAGGAAGGAGGAGGAAGGAAGAA	28607
Qy	1539	gggatatgtagtgactgctttggagctcgggctggttgcgtgtgcttgccagaaaagtccag	1598
Db	28606	GCAAGGAGGGGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	28547
Qy	1599	gctaagatctcatcggctcttggcttggggcccttggcaggttgtgtgaccccttggtctgg	1658
Db	28546	GCAAGGAAGGAAGGAAGCAAGGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG	28487
Qy	1659	acagggaaccagagagcagcagcactcgcgcagatgggagggcagtggtgtctgtg	1718
Db	28486	ACAGAAAGAAAGGAAGAAAGAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	28427
Qy	1719	gatatgtggccaggttcagttgggaagctgaaggatgagcagaccttaggtctcaggaagga	1778

Db	28426	GAGGAGGGAAGGAAAGGAAGGAGGGAAGGAGGAAGGAGGAACAAGGGA	28367
Qy	1779	gggctgcctgaagtgggggcacatcacctgaccagaaaggaaaaactggcagtg-----	1833
Db	28366	GGGAGGAGGGAAGACGGGAGAGAGAGAGAAAGAAAGAAAGAAAGGAAGGAA	28307
Qy	1834	-----ccagggtctggaaggggcctgcattgagcttgaaaaaaactataataga	1881
Db	28306	GGAAGGAAGAAAGGAAGGAAGGAAAATCTAAGAGAATAGAAACTCAGGATAATGAT	28247
Qy	1882	attggttaccaattttattattattattattattattattattatttttttttgagataagatct	1941
Db	28246	TTTGGTCTTGGGGTTTTTTTGTAGCTTTATTTCTTAAATTTCTTTTGAGATGAAGTCT	28187
Qy	1942	cactcccttgcgaaggtcgagtcggtggtgcttatctcagctcactgcgaacctctgcct	2001
Db	28186	TCTCTGTCAACCCAGGCTGGAGTCAGTGGCGACAATCTCGGCTCATTTGCAACCTCCACCT	28127
Qy	2002	ccaggatacaagtgtattctccagcctcagcctccccaggtagctgggattacaagcatgc	2061
Db	28126	CCCAGGTTCAAGCAATTTCTTGTCTCAGCCTCTCCA-GTAACTGGGACTACAGGTTTCAC	28068
Qy	2062	accacatgcctcgatgaattttttttttttttttttttttttttttttttccacaggttgccc	2121
Db	28067	GCCACCAAGGCCAGCATATTTTGTATTTTAAATAGAGATGGGATTTTACCATATTGGTC	28008
Qy	2122	agactggtctgaactctcagctcagtgatctgcctgctcgctgcctcccaagtgctg	2181
Db	28007	AGGCTGGTCTCGAATCTGTACCCAGAGTGATCTACCCACCTCGGCTCCCAAGTGCTG	27948
Qy	2182	gaattacagatgtgagccactgtccctggcctggtgtaccacacatttataaatggagtgat	2241
Db	27947	GGATTACAGGCATGAGCCACTGTGCTTGGC-----	27916
Qy	2242	ttcaaccttttatgtggaattacagctgtgttttttttttttttttttttttttttttttttt	2301
Db	27917	-----CATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGACAGAGTCTTGC	27880
Qy	2302	tctgtccccaggctgaagtgcagtaatgcaatctcagctcactgcgaaccttagcctcct	2361
Db	27879	TCTGTCAACCCAGGCTGGGTCGAATGGCACCATCTCAGCTCAACAACCTCTGCCCTCCC	27820
Qy	2362	gggttcaagcaattctcctgcctcagccacactgagtagcctggggttacagggcatgcacc	2421
Db	27819	AGGTTCAAGCAATTTCTCTGCTTCTCAGCCTCTCAAGTAA-CTGGGACTACAGCGCATGCC	27761
Qy	2422	accacgcaggctaattt	2481
Db	27760	ACCACGCCTAGCTAA-TTTTTGTATTTTAGTAGAGCGAGGGTTTTCACCATGTTTGGCGAG	27702
Qy	2482	gctggtctgaactcctcagctcaggtgatccgcccttgccctcccaagtgctcagg	2541
Db	27701	GTTGGTCTTGAACTCCTGACCTCAAAATGATCCGCTCCCTCTGCTTCCCAAGTGCTGGG	27642
Qy	2542	attacagtggaacacacttgccacgctggtgctatcgttt	2584
Db	27641	ATTCAGGCATGAACCACTGCGCGGGGCTTTTTTTTTTTTTTTTTTTTTTTT	27599
RESULT	3		
ID	AAF97846/c		
XX	AAF97846 standard; DNA; 51474 BP.		
AC	AAF97846;		
XX			
DT	31-MAY-2001 (first entry)		
XX			
DE	Human neuroblastoma cell line NB-1 lp36 nucleotide sequence SEQ ID NO: 6		
XX			
KW	Human; chromosome 1; lp36; neuroblastoma cell line; NB-1; anticancer;		
KW	tumour suppressor; human lp36 homozygosity deletion domain; tumour;		
KW	diagnosis; ds.		

CC illness. The method distinguishes neuropsychiatric disorders from
CC neurological disorders, which enables more accurate evaluation and
CC prescription of medical treatment. The present sequence represents the
CC human ysl cDNA sequence.
XX
SQ Sequence 119950 BP; 34471 A; 23730 C; 24660 G; 37033 T; 56 other;

Query Match 11.4%; Score 382.8; DB 20; Length 119950;
Best Local Similarity 76.4%; Pred. No. 9.7e-78;
Matches 547; Conservative 0; Mismatches 162; Indels 7; Gaps 6;

Qy 1871 actataatagaattgggtaccattttattttattttattttattttattttttttttg 1930
Dy 106332 acagtatgtaattaccattggaatttttttttttttttttttttttttttttt 106391

Qy 1931 agata-gagtcctactcctgtcctaagctgagctgagtgctgctatctcagctcag 1989
Dy 106392 agacaagagctcctgttctgaccacagctgagtgagtgctgctatctcagctcag 106451

Qy 1990 caactctgctccacagatcaagtatttccagctcagctcagctcagctgagtgagga 2049
Dy 106452 caactcgcctccaggttcaagtatttccgctcagctcagctcagctgagga 106510

Qy 2050 ttacaagcatg-caccacacatgctgggataatttttttttttttttttttttt 2108
Dy 106511 ttacagtggtgcccacatgcccagttagtttttttttttttttttttttttt 106570

Qy 2109 caccaggtgcccagactggtctcgaaactctgacctcaggtgactcctgctcagcc 2168
Dy 106571 cgcctatgtgcccagctggtctcaaacctcagctcaggtgactcagcc 106629

Qy 2169 tcccaaatgctggaattacagatgtgagccactgctcctgctggttaccacatttt 2228
Dy 106630 tcccaaatgctggaattacagatgtgaaacacacatgcccagctggtgca 106689

Qy 2229 aaaatggagtgatttcaaccttttatgtggatttacagctgttttttttttttt 2288
Dy 106690 atagtctaatatcagaaaattctgaagtggaagtgttttttttttttttt 106749

Qy 2289 gacaaagtctggtctgtcaccacagctgagtgagtgagtgagtgagtgagtgag 2348
Dy 106750 gatggagtcaccactctgtcaccacagctgagtgagtgagtgagtgagtgag 106809

Qy 2349 acctgactcctggttccaaagcaattctcgtcctcagccacactgagtgagtgag 2408
Dy 106810 acctcaccctccaggttccaaagcaattctcgtcctcagccacactgagtgag 106868

Qy 2409 acaggatgcaccacacacagcagctaattttttttttttttttttttttttttt 2468
Dy 106869 acagatgcccacacacagcagctaattttttttttttttttttttttttttt 106928

Qy 2469 ccattgtgcccagctggtctcgaaactcctgacctcaggtgagtgagtgagtgag 2528
Dy 106929 cca-tgtgcccagagtggtctcgaaactcctgacctc--gtgatgctgctcag 106986

Qy 2529 ccaaaagtctaggattacagtgagggaacacacttgcacagcctgtggctatcggtt 2584
Dy 106987 ccaaaagtctaggattacagcattagccacagcagccagcagcagcagcagc 107042

RESULT 5
AAH98793
ID AAH98793 standard; cDNA; 1624 BP.
XX
AC AAH98793;
XX
DT 12-OCT-2001 (first entry)
XX
DE Human EST-derived coding sequence SEQ ID NO: 650.
XX
KW Human; sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;
KW tomato; monkey; dog; sea urchin; expressed sequence tag; EST;

KW diagnostics; forensic test; gene mapping; genetic disorder;
KW biodiversity; gene therapy; nutrition; ss.
XX
OS Homo sapiens.
XX
PN WO200154477-A2.
XX
PD 02-AUG-2001.
XX
PF 25-JAN-2001; 2001WO-US02687.
XX
PR 25-JAN-2000; 2000US-0491404.
PR 17-JUL-2000; 2000US-0617746.
PR 03-AUG-2000; 2000US-0631451.
PR 15-SEP-2000; 2000US-0663870.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V;
PI Cao Y, Drmanac RA, Zhang J, Werhman T;
XX
DR WPI: 2001-476164/51.
DR P-PSDB; AAM24134.
XX
PT Isolated polypeptide for treatment of diseases, diagnostics, raising
PT antibodies and research use -
XX
PS Claim 1; Page 615; 1275pp; English.
XX
CC The present invention provides the protein and coding sequences of novel
CC proteins from a variety of organisms, including human, dog, cat, horse,
CC cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea
CC urchin and tomato. These were derived from expressed sequence tags (ESTs)
CC from the organism of interest. They can be used in diagnostics,
CC forensics, gene mapping, identification of mutations, to assess
CC biodiversity and for nutritional purposes. The present sequence is a cDNA
CC of the invention.
XX
SQ Sequence 1624 BP; 426 A; 347 C; 343 G; 507 T; 1 other;

Query Match 11.2%; Score 375.2; DB 22; Length 1624;
Best Local Similarity 76.2%; Pred. No. 1.2e-76;
Matches 532; Conservative 0; Mismatches 148; Indels 18; Gaps 5;

Qy 1875 taatagaattggtaccattttattttattttattttattttattttttttttttgagat 1934
Dy 76 ttattcacatgtgatactaaccttgaaggttttttttttttttttttttttttttttt 135

Qy 1935 agagtctcactc-ccttgctaagctgagtgagtcggtgctatctcagctcactgcaac 1993
Dy 136 ggagtctctctctgcccagctgagtgagtgagtgagtgagtgagtgagtgagtgag 195

Qy 1994 ctctgctcccaggaatcaagtgattctccagctcagctcagctcagctcagctcagct 2053
Dy 196 ctccacctcccaggtcctcagctcagctcagctcagctcagctcagctcagctcagct 254

Qy 2054 aagcatgcaccacatgctggagataatttttttttttttttttttttttttttttt 2113
Dy 255 agacactctccacacacccggcctaattttatactttcggcagagaggggtttccacca 314

Qy 2114 ggttgccagactggtctcgaactctcagctcaggtgagctgctgctgctgctgctgct 2173
Dy 315 tgttgaccaggtgctgctcgaactcagctcagctcagctcagctcagctcagctcagct 374

Qy 2174 aagtctggaattacagatgtgagcactgtccctggtcgtggttaccacacatttttaaaat 2233
Dy 375 aagtctggtgattacagatgtgagcaccatgctcgtggtcgtggtt-----tt 420

Qy 2234 ggagtgtttaccacctttatgttgatttacagctgttttttttttttttttttttttt 2293
Dy 421 gttttctgttt 480

Db 17653 actctgacctcagatgatcaaccgccttggtgcctccgaagtctgggattataggcgt 17712

Qy 2553 gaaccacttgcagcctgtgtctatcgttttaaac 2588
 || ||||| || | | | | | | | |

Db 17713 gagccacactcctgtcactcctctctctgtac 17748
 : :

RESULT 14

AA02697

ID AAD02697 standard; DNA: 160552 BP.

XX AAD02697:

AC AAD02697:

XX

XX

DT 02-MAY-2001 (first entry)

XX Human glycosyl sulfotransferase-4 (GST-4) genomic DNA.

DE

XX

KW Human; glycosyl sulfotransferase-4; GST-4; immunosuppressive;
 KW therapy; selectin binding inhibitor; gene therapy; inflammation;
 KW systemic lupus erythematosus; SLE; rheumatoid arthritis; diabetes;
 KW polyarteritis nodosa; polymyositis; systemic sclerosis; dermatitis;
 KW glomerulonephritis; myasthenia gravis; Sjogren's syndrome; adenolitis;
 KW Hashimoto's disease; Grave's disease; hypoparathyroidism; anaemia;
 KW demyelinating disease; cirrhosis; ulcerative colitis; allergic rhinitis;
 KW myocarditis; adult respiratory distress syndrome; eczema; psoriasis;
 KW asthma; hypersensitivity; rheumatic fever; tissue rejection;
 KW chromosome 16q23.1; ds.

XX

OS Homo sapiens.

XX

Key	Location/Qualifiers
FT exon	32847..32922
FT	/*tag= a
FT	/number= 1
FT intron	/label= 4a_5U4
FT	32923..35592
FT	/*tag= b
FT	/cons_splice= (5'site:NO, 3'site:YES)
FT exon	35593..35674
FT	/*tag= c
FT	/number= 2
FT intron	/label= 4a_5U3
FT	35675..45093
FT	/*tag= d
FT exon	45094..45185
FT	/*tag= e
FT	/number= 3
FT intron	/label= 4a_5U2
FT	45186..46633
FT	/*tag= f
FT exon	/cons_splice= (5'site:NO, 3'site:NO)
FT	46634..46700
FT	/*tag= g
FT	/number= 4
FT intron	/label= 4a_5U1
FT	46701..47938
FT	/*tag= h
FT exon	/cons_splice= (5'site:YES, 3'site:NO)
FT	47939..49746
FT	/*tag= i
FT	/number= 5
FT 5'UTR	/note= "Includes 17 base pairs of 5'UTR, the ORF and all of 3'UTR"
FT	47939..47955
FT	/*tag= j
FT	/note= "Portion of 5' untranslated region (5'UTR)"
FT	47956..49128
FT	/*tag= k
FT	/product= "Human glycosyl transferase-4alpha (GST-4alpha)"
FT 3'UTR	49129..49746
FT	/*tag= l
FT exon	83257..83347

FT

FT intron

FT

FT exon

FT

FT intron

FT

FT exon

FT

FT 5'UTR

FT

FT CDS

FT

FT 3'UTR

FT

XX WO200106015-A1.

XX

XX 25-JAN-2001.

XX

PF 19-JUL-2000; 2000WO-US19741.

XX

XX 20-JUL-1999; 99US-0144694.

PR

PR 13-JUL-2000; 2000US-0593828.

XX

XX (REGC) UNIV CALIFORNIA.

XX

PI Rosen SD, Lee JK, Hemmerich S;

XX

XX WPI: 2001-138471/14.

DR

DR P-PSDB; AAY72639, AAY72640.

XX

PT New glycosyl sulfotransferases (GST)-4alpha, GST-4beta and GST-6 for
 diagnostic and therapeutic agent screening applications -

XX

PS Example 1; Page 62-104; 128pp; English.

XX

CC The present sequence is human glycosyl sulfotransferase-4 (GST-4) genomic
 CC DNA encoding GST-4alpha and GST-4beta. GST-4 gene is found on
 CC chromosome 16q23.1.

CC

CC GST is a type 2 membrane protein useful for inhibiting a binding event
 CC between a selectin and a selectin ligand, which comprises contacting the
 CC selectin with a non-sulphated selectin ligand, GST and a small molecular
 CC agent that inhibits the sulphation activity of GST. GST is also useful
 CC in inhibiting a selectin mediated binding event. GST is useful in gene
 CC therapy to treat disorders such as acute or chronic inflammation,
 CC systemic lupus erythematosus (SLE), rheumatoid arthritis, polyarteritis
 CC nodosa, polymyositis, dermatomyositis, systemic sclerosis, diabetes,
 CC glomerulonephritis, myasthenia gravis, Sjogren's syndrome, Hashimoto's
 CC disease, Grave's disease, adenolitis, hypoparathyroidism, pernicious
 CC anaemia, demyelinating diseases, cirrhosis, ulcerative colitis,
 CC dermatitis, myocarditis, regional enteritis, adult respiratory distress
 CC syndrome, infantile eczema, psoriasis lichen planus, allergic rhinitis,
 CC bronchial asthma, hypersensitivity, rheumatic fever and tissue rejection
 CC during transplantation.

XX

SQ Sequence 160552 BP; 40281 A; 37573 C; 38015 G; 44564 T; 119 other:

Query Match 10.8%; Score 363.6; DB 22; Length 160552;
 Best Local Similarity 53.9%; Pred. No. 2.8e-73;
 Matches 1401; Conservative 2; Mismatches 991; Indels 204; Gaps 24;

Qy 2 aaaggtcaggcactgtggtctatcgtctaatccccagcactttggaggcgagcggtt 61


```
QY 2160 gctcggcctcccaagtgctggaattacagatgtgagccactgtccctggcctggttac 2219
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 140311 cctcagcctctaaagtgtggtgattacagcggtgagccactacgctggcctgagttg 140370
QY 2220 ccacattttaaatggagtgattccacccttttatgtgg----attacagctgtgtttt 2275
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 140371 tgcacttaaatgggtacattttatgtgtgtgaaattgtgtcctaataaggctttttttt 140430
QY 2276 ttttttttttgagacaaagtctggtcctgtccaccagcggtgagtgcaagtaaatgcaatc 2335
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 140431 ttttttttttgagcggagtgctcctctgtccaccagcgcgagtgcaagtgcgcgatc 140490
QY 2336 tcaagtcactgcaaccttagcctctctggtgttcaagcaattctcctgcagccacctga 2395
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 140491 tcgggtcactgcaagctcgcctccctcaggttcattgccattctccggtcagctcccca 140550
QY 2396 gtacgtcgtgggttacagcgtatgcaccacacgagcaggttaatttttttta-ttttttagta 2454
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 140551 gtgg-ctgggactacagcgctcgtccaccagcccccggtcaatttttatatttttagta 140609
QY 2455 gagatggggtttcgcacatgttgccagcgctgtctcgaactcctgacctcaggtgatccg 2514
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 140610 cagacgggggtttcacctggttagccagatggtctcgatcctctgaacct--tgtgatccg 140667
QY 2515 cccgccttggtcctcccaa 2532
||||| ||||| |||||
Db 140668 cctgcctcagcctcccaa 140685

RESULT 15
AAS11614/c
ID AAS11614 standard; DNA; 236303 BP.
XX
AC AAS11614;
XX
DT 24-OCT-2001 (first entry)
XX
DE Human genomic DNA containing exons 2-17 of the CRIM1 gene.
XX
KW CRIM-1; Human; human chromosome 2p21-16.3; ophthalmological;
neuroprotective; renal; osteopathic; dental; vulnery; immunogen;
KW antibody; gene therapy; neurodegenerative disease; eye disorder;
KW cataract; bone morphogenic protein; BMG; renal disease; bone abnormality;
KW tooth abnormality; wound; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT exon /*tag= a
FT /*number= 2
FT intron 33278..77746
FT /*tag= b
FT /*number= 2
FT exon 77747..77989
FT /*tag= c
FT /*number= 3
FT intron 77990..79103
FT /*tag= d
FT /*number= 3
FT exon 79104..79224
FT /*tag= e
FT /*number= 4
FT intron 79225..101022
FT /*tag= f
FT /*number= 4
FT exon 101023..101144
FT /*tag= g
FT /*number= 5
FT intron 101145..113377
FT /*tag= h
FT /*number= 5
FT exon 113378..113560
FT /*tag= i
FT /*number= 6
FT intron 113561..115985
FT /*tag= j
FT /*number= 6
FT exon 115986..116183
FT /*tag= k
FT /*number= 7
FT intron 116184..135707
FT /*tag= l
FT /*number= 7
FT exon 135708..135836
FT /*tag= m
FT /*number= 8
FT intron 135837..146471
FT /*tag= n
FT /*number= 8
FT exon 146472..146628
FT /*tag= o
FT /*number= 9
FT intron 146629..148761
FT /*tag= p
FT /*number= 9
FT exon 148762..148883
FT /*tag= q
FT /*number= 10
FT intron 148884..150044
FT /*tag= r
FT /*number= 10
FT exon 150045..150254
FT /*tag= s
FT /*number= 11
FT intron 150255..153815
FT /*tag= t
FT /*number= 11
FT exon 153816..154031
FT /*tag= u
FT /*number= 12
FT intron 154032..158580
FT /*tag= v
FT /*number= 12
FT exon 158581..158802
FT /*tag= w
FT /*number= 13
FT intron 158803..173982
FT /*tag= x
FT /*number= 13
FT exon 173983..174177
FT /*tag= y
FT /*number= 14
FT intron 174178..181006
FT /*tag= z
FT /*number= 14
FT exon 181007..181129
FT /*tag= aa
FT /*number= 15
FT intron 181130..183612
FT /*tag= ab
FT /*number= 15
FT exon 183613..183800
FT /*tag= ac
FT /*number= 16
FT intron 183801..185152
FT /*tag= ad
FT /*number= 16
FT exon 185153..187765
FT /*tag= ae
FT /*number= 17
XX
PN WO200138519-A1.
XX
PD 31-MAY-2001.
```


THIS PAGE BLANK (USPTO)

Parent case 10/767341

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 11, 2002, 16:31:36 ; Search time 14958.6 Seconds
(without alignments)
3706.691 Million cell updates/sec

Title: US-09-820-005-3_COPY_10200_13560
Perfect score: 3361
Sequence: 1 aaagtcagcagctgtggc.....ctgtccaagtgggtgcatt 3361

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

- 1: gb.ba.*
- 2: gb.htg.*
- 3: gb.in.*
- 4: gb.om.*
- 5: gb.ov.*
- 6: gb.pat.*
- 7: gb.ph.*
- 8: gb.pl.*
- 9: gb.pr.*
- 10: gb.ro.*
- 11: gb.sts.*
- 12: gb.sy.*
- 13: gb.un.*
- 14: gb.vi.*
- 15: em.ba.*
- 16: em.fun.*
- 17: em.hum.*
- 18: em.in.*
- 19: em.om.*
- 20: em.or.*
- 21: em.ov.*
- 22: em.pat.*
- 23: em.ph.*
- 24: em.pl.*
- 25: em.ro.*
- 26: em.sts.*
- 27: em.sy.*
- 28: em.un.*
- 29: em.vi.*
- 30: em.htgo_hum.*
- 31: em.htgo_inv.*
- 32: em.htgo_rod.*
- 33: em.htg_hum.*
- 34: em.htg_inv.*
- 35: em.htg_rod.*
- 36: em.htg_other.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	3312.8	98.6	124526	9	AC005080	AC005080 Homo sapi
2	3306.6	98.4	230552	9	AC005098	AC005098 Homo sapi
3	3304.2	98.3	131359	9	AC004883	AC004883 Homo sapi
4	3299.4	98.2	169604	9	AC008384	AC008384 Homo sapi
5	3293.2	98.0	275197	9	AC004166	AC004166 Homo sapi
6	3254.2	96.8	17302	9	AF184614	AF184614 Homo sapi
7	2993.2	89.1	3348	9	HS47P43S02	HS47P43S02
8	2990	89.0	3350	9	HS47P40S02	HS47P40S02
9	2984.4	88.8	3349	9	HS47P40X02	HS47P40X02
10	2845.8	84.7	3200	9	HS47P41S03	HS47P41S03
11	2823.6	84.0	2899	9	HS47L14S02	HS47L14S02
c 12	2668.8	79.4	184558	2	AC092405	AC092405 Papio cyn
c 13	1719.8	51.2	149830	2	AC027219	AC027219 Homo sapi
c 14	886.4	26.4	918	9	HSNCF1S4	HSNCF1S4
15	870.8	25.9	904	9	HS39NCF5	HS39NCF5
16	772	23.0	799	9	HS39NCF6	HS39NCF6
17	505.8	15.0	124095	9	AC002477	AC002477 Human PAC
18	500.8	14.9	149830	2	AC027219	AC027219 Homo sapi
19	497.2	14.8	149490	2	AL589985	AL589985 Homo sapi
c 20	489.6	14.6	6901	9	HS694E4	HS694E4
21	482.6	14.4	315835	2	AL591212	AL591212 Homo sapi
22	481	14.3	178451	9	AL139396	AL139396 Human DNA
23	476.8	14.2	188833	9	HS268H5	HS268H5
c 24	472.4	14.1	178451	9	AL139396	AL139396 Human DNA
c 25	464.8	13.8	191717	2	AC012482	AC012482 Homo sapi
26	464.4	13.8	180675	2	AL354671	AL354671 Homo sapi
27	464.4	13.8	240501	2	AL590226	AL590226 Homo sapi
c 28	459.6	13.7	123829	9	HSJ459A13	HSJ459A13
29	459.6	13.7	181386	9	AC007686	AC007686 Homo sapi
30	458.6	13.6	115079	2	AC008105	AC008105 Homo sapi
31	457.8	13.6	107808	2	AL162271	AL162271 Homo sapi
32	457.8	13.6	213432	2	AC068198	AC068198 Homo sapi
c 33	456.4	13.6	145833	2	AC026539	AC026539 Homo sapi
c 34	456.2	13.6	163035	2	AC006405	AC006405 Homo sapi
c 35	456.2	13.6	180407	2	AL590822	AL590822 Homo sapi
c 36	455	13.5	166549	2	AC025990	AC025990 Homo sapi
c 37	455	13.5	183996	9	AF196779	AF196779 Homo sapi
c 38	454.8	13.5	240455	2	AC010290	AC010290 Homo sapi
39	454	13.5	120955	9	HUAC002310	AC002310 Human Chr
c 40	452.8	13.5	167515	2	AC091864	AC091864 Homo sapi
c 41	452.2	13.5	120955	9	HUAC002310	AC002310 Human Chr
42	452.2	13.5	162365	2	AC018521	AC018521 Homo sapi
c 43	448.6	13.3	180236	2	AC020600	AC020600 Homo sapi
c 44	448.2	13.3	208774	2	AC009634	AC009634 Homo sapi
45	447.8	13.3	99876	9	HSDJ90108	AL078461 Human DNA

ALIGNMENTS

RESULT	1
AC005080	
LOCUS	AC005080 124526 bp DNA PRI 02-OCT-2000
DEFINITION	Homo sapiens BAC clone CTA-269p13 from 7q11.2, complete sequence.
ACCESSION	AC005080
VERSION	AC005080.2 GI:7770715
KEYWORDS	HTG.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
TITLE	1 (bases 1 to 124526)
JOURNAL	Sulston,J.E. and Waterston,R.
MEDLINE	Toward a complete human genome sequence
REFERENCE	Genome Res. 8 (11), 1097-1108 (1998)
AUTHORS	99063792
TITLE	2 (bases 1 to 124526)
JOURNAL	Scott,K., Layman,D., Kalicki,J. and Harmon,G.
	The sequence of Homo sapiens BAC clone CTA-269p13
	Unpublished

```

REFERENCE
AUTHORS
TITLE
JOURNAL
3 (bases 1 to 124526)
Waterston,R.H.
Direct Submission
Submitted (12-JUN-1998) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
4 (bases 1 to 124526)
Waterston,R.H.
Direct Submission
Submitted (12-MAY-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
5 (bases 1 to 124526)
Waterston,R.H.
Direct Submission
Submitted (12-JUN-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
6 (bases 1 to 124526)
Waterston,R.
Direct Submission
Submitted (02-OCT-2000) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On May 12, 2000 this sequence version replaced gi:3212911.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: sapiens@watson.wustl.edu
----- Summary Statistics
----- Center project name: H_RG269P13
-----

```

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/GRB/CHR7>, send <mailto:egreen@nhgri.nih.gov>, or see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

Clone CTA-269P13 is from a release of the human BAC library CTRB-HS-A. The library contains cloned DNA from human sperm. See: Shizuya et al., Proc. Natl. Acad. Sci. USA 89:8794-7 (1992); U-J. Kim et al., Genomics 34:213-8 (1996). The clone is available from Research Genetics, Inc. (<http://www.resgen.com>).
VECTOR: pBelOBAC11
Selection: chloramphenicol

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP4-754G14; the clone sequenced to the right is RP11-396K3. Actual start of this clone is at base position 1 of CTA-269P13; actual end is at base position 124526 of CTA-269P13.

The clone CTA-269P13 may contain a transposon in the growth of the clone, which is not part of the submitted sequence.

FEATURES

Source

```

Location/Qualifiers
1..124526
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="7"
/map="7q11.2"
/clone="CTA-269P13"
/clone_lib="CTB-HS-A"
1..138
/note="match to EST AA643257 (NID:g2568475) nr61h09.sl"
1..338
/note="match to EST D44965 (NID:gl572440)"
42..118
/note="similar to EST D44969 (NID:gl572444)"
complement(106..396)
/db_xref="GI:3800554"
119..483
/note="similar to EST AW236834 (NID:g6569223) xm49f08.xl"
119..483
/note="similar to EST AI609282 (NID:g4618449) tw83h06.xl"
119..378
/note="similar to EST D44969 (NID:gl572444)"
146..483
/note="similar to EST N80671 (NID:gl243372) zb03a07.sl"
164..483
/note="similar to EST AW204636 (NID:g6504108)"
204..206
/note="match to EST AA643257 (NID:g2568475) nr61h09.sl"
306..652
/note="match to EST AA804480 (NID:g2875993) nz33c09.sl"
335..520
/note="similar to EST AW177547 (NID:g6443584)"
577..666
/note="similar to EST AW177547 (NID:g6443584)"
595..711
/rpt_family="Alu"
712..740
/rpt_family="(CAAAA)n"
741..898
/rpt_family="Alu"
899..919
/rpt_family="AT-rich"
965..1106
/rpt_family="Alu"
1107..1395
/rpt_family="Alu"
1396..1570
/rpt_family="Alu"
1571..1662
/rpt_family="L1"
1665..1817
/rpt_family="L1"
1834..1886
/rpt_family="L2"
1888..2189
/rpt_family="Alu"
2256..2546
/rpt_family="Alu"
3070..3391
/rpt_family="Alu"
3493..3780
/rpt_family="Alu"
3792..4101
/rpt_family="Alu"
4108..4232
/rpt_family="Alu"
4247..4422
/rpt_family="Alu"
4423..4729
/rpt_family="Alu"
4730..4866
/rpt_family="Alu"
4941..5099

```

misc_feature

misc_feature

misc_feature

STS

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

```
repeat_region /rpt_family="L1"
5222..5527 /rpt_family="Alu"
repeat_region 5812..5883 /rpt_family="Alu"
repeat_region 5884..6194 /rpt_family="MER1_type"
repeat_region 6207..6397 /rpt_family="Alu"
repeat_region 6398..6469 /rpt_family="MaLR"
repeat_region 6470..6771 /rpt_family="MaLR"
repeat_region 7423..7468 /rpt_family="Alu"
repeat_region 7469..7644 /rpt_family="MaLR"
repeat_region 7645..7774 /rpt_family="MaLR"
repeat_region 7860..8155 /rpt_family="MER1_type"
repeat_region 8194..8466 /rpt_family="Alu"
repeat_region 8479..8599 /rpt_family="Alu"
repeat_region 8599..8600 /rpt_family="MIR"
misc_feature /note="similar to EST W78139 (NID:g273876)"
8601..8916 /rpt_family="Alu"
8787..8803 /rpt_family="Alu"
misc_feature /note="similar to EST AA463503 (NID:g2188387) zx72c10.s1"
8810..9112

Query Match 98.6%; Score 3312.8; DB 9; Length 124526;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 3351; Conservative 0; Mismatches 7; Indels 4; Gaps 3;

Qy 1 aaaaagtcaggcaactgtggctcaatccgttaataccagcaactttggaggccagagcggt 60
Db 116133 AAAAGGTCAGGCACATGTGGCTCATCCCTGTAATCCAGCACATTTGGAGGCGCGAGCGGT 116192

Qy 61 tgaatcaactaggtcagaggttcagaccagctgaccaatatatcgtaaaactccatctc 120
Db 116193 TGAATCACTGAGGTCAGAGGTTCAGACCAGCCCTGACCAATATCGTAAACTCCATCTC 116252

Qy 121 tactaaaaatacaaaaattagccaggtgtggtggcggtgctttagtcccagctacttg 180
Db 116253 TACTAAAAATACAAAATTAGCCAGGTGTGGTGGCAGGTGCTTGTAGTCCCAAGCTACTTG 116312

Qy 181 ggagctcaggcaggtgaattacttgaacctggagggtggaaggttgcgaatgagccaaagt 240
Db 116313 GGAGGCTTCAGGCAGGTGAATTAATTAACCTGGGAGGTGGAGGTTGCAATGAGCCAAAGT 116372

Qy 241 tgcacctgcactccagtgcagagcagactccatctc-aaaaaaaaaaaaaaaaa 299
Db 116373 TGCACCACTGCACACTCCAGTGACAGAGCAGACTCCATCTCAAAAAAATAAAAAA 116432

Qy 300 gttggaaaagccagggtgagtcgctccacgcctgtaatcccaacactttaagaggctg 359
Db 116433 GTTGGAAAAGCCAGGTGCAGTGGCTCCACGCCTGTAATCCCAACACTTTTAAGAGGCTG 116492

Qy 360 aggtgggaaatccctttgagccagaggttcagaccagcctgggcatgtgtcccaagacc 419
Db 116493 AGGTGGGAAATCCTTTGAGCCAGGAGTTTCAGACACAGCCTTGTCCCAAGACC 116552

Qy 420 ttgtctttacaaaattagccgggtgtgtggaatacgtctgtgtgtcccaactattcgg 479
Db 116553 TTGTCTTTACAAAAAATTAGCCGGGTGTGGTGGCATACGTCTGTGTGCCCACTATTGG 116612
```

|||||
Db 117693 GGAGCTGGCTGGTTCCTGGCTTGGCAGAAAGTCAGGGCTTAAGATCAICGTCCTCG 117752
QY 1620 gtttggggccctgcaggttctgtagcccttggctctgagacaggaagagagagag 1679
Db 117753 GCTTGGGGCCCTGGCAGGTTGTGATGGCTTGGCTGGACAGGAACAGAGAGAGAG 117812
QY 1680 cagacactcggaagtgagagagccagtggtctgtagatatgtgccaggttcaatg 1739
Db 117813 CAGACACTCGGAGAGTGGGAGGCCAGTGGTGTCTGTGGATATGTGGCCAGGTTCAAGTG 117872
QY 1740 ggaagctgaagatgagcagacacttagctcaggaagagggctgcctggaagtggggc 1799
Db 117873 GGAAGCTGAAGGATGAGCAGACCTTAGGCTCAGGAAGAGGGCTGCCCTGGAAGTGGGGC 117932
QY 1800 atcaactgaccagaaaggggaaaaactggcagdgccagggctggatggggcctgcatga 1859
Db 117933 ATCATCACTGACCAAGAAAGGAAAACTGGCAGCTGCCAGGGCTGGATGGGGCCGTCAATTGA 117992
QY 1860 gcttaaaaaaactataatagaaattggttaccattttttattattattattattattatt 1919
Db 117993 GCTTGAaaaaaactataatagaaattggttaccattttttattattattattattattatt 118052
QY 1920 ttacttttttgatagagctcactccctctgctaaggctggagtgccggtgctgctatct 1979
Db 118053 TTACTTTTTTGGATAGAGTCTCACTCCCTTCTAAGGCTGGAGTGGGTGTCTATCT 118112
QY 1980 cagctcaactgaaacctctgcctcccaaggatcaagtgaattctccagcctcagcctcccaag 2039
Db 118113 CAGCTCACTGCAACCTCTGCCCTCCAGGATCAAGTCAATCTCCAGCCTCAGCCTCCCCAG 118172
QY 2040 gtagctgggattacaagcatgacacacacatgcttgataatttttattttttttagtgag 2099
Db 118173 GTAGCTGGGATTACAAGCATGACACCATGCTGGATPAATTTTTGTATTTTAGTTGAG 118232
QY 2100 acggggtttccacaggttggccagactgctctcgaaactctgaacctcgaactcgaatgctgct 2159
Db 118233 ACGGGGTTTTACAGAGTTGGCCAGAGCTGCTCTCGAACTTCTGACCTCAGTGATCTGCT 118292
QY 2160 gctcgccctcccaagtgctggaattacagatgagccactgctccctggcctggttac 2219
Db 118293 GCTTGGCCCTCCCAAGTCTGGAATACAGATGTAGGCCACTGTCCTGGCCCTGGTTAC 118352
QY 2220 ccacatttttaaaatggagtgatttccaccttttattgtgatttaccagctgttttttttt 2279
Db 118353 CCACATTTTAAATGGAGTGATTTACACCTTTTATGTGGATTTACAGCT - TTTTITTTTT 118410
QY 2280 ttttttgagacaaagtctgctctgtccacagcgtcgagtgcaatgcaatctcag 2339
Db 118411 TTTTGTGAGCAAAAGTCTGGCTCTGTCAACCAGGCTGGAGTGCAGTAATGCAATCTCAG 118470
QY 2340 ctcaactgcaaccttagctctcctgggttcaagcaattctcctccctcagccactgagtag 2399
Db 118471 CTCACCTGCAACCTTAGCTCTGGGTTCAAGCAATTCCTCGCTCAGCCACTGAGTAG 118530
QY 2400 cctggggttacaggcatgacaccacacagccagggctaatttttttatttttttagtagagat 2459
Db 118531 -CTGGGATTACAGGCATGACACCAACCCAGCCAGGCTAAATTTTGTATTTTGTAGTAGAGAT 118589
QY 2460 ggggttcccoactgttggccagcgtggtctcgaaactcctgaactcctcagctcagtcgcccgcc 2519
Db 118590 GGGGTTTCCGCCATGTTGGCCAGGCTGGTCTCGAACTCTCGAACTCCTGAGCTCAGGTTGATCCGCCGCC 118649
QY 2520 ctggcctcccaaaagtctgagattacaggttggaacacacctgcccacctgtggtctat 2579
Db 118650 CTGGCCCTCCCAAAAGTCTAGGATTAACAGTGGGAACCACTCGCCACCCCTGTGCTAT 118709
QY 2580 cgtttaaacactgggaagcctgcagcccccagggccgacagtttagctcagctgagcagct 2639
Db 118710 CGTTTAAACACTGGGAAGCCCTGCAGCCCCCAGCCGACAGTTAGCTGCAGCTGAGCACT 118769
QY 2640 tccacgtgcaggtagacaggtgctccacccacctactcatggtgctgctctgtagtag 2699
|||||

Db 118770 TCCCAGTGCCAGGTAGACGGATGCTCCACCCACCTACTCATGGTGTATCTCTTGTCTATAG 118829
QY 2700 tgaagtgtctggacagacacttcaatcgttatgggattctctgtccccagagtggttgagcaa 2759
Db 118830 TGAAGTGTCTGGACAGACTTTCATGCTTATGGGATCTCTGTGCTCCAGAGTGGTGGCAA 118889
QY 2760 tgaatggagtggaacaagctcacctgggtgtagggggcagagggccgaagtccagagtgt 2819
Db 118890 TGAATGGGAGTGGACAGCTCACCTGGGTGTAGGGGGCAGAGGGCCGAAAGTCCACAGTGT 118949
QY 2820 acccccagagtggtgtgcagcagaggcttgcagagggatctgggagtggagcagggaggtg 2879
Db 118950 ACCCCACAGTGGTGTCCAGCAGGAGCTTGCAGGGATCTGGGATGGAGCAGGAGGGTG 119009
QY 2880 gaggagagagaccccaagaagaggggaaactgtgggcccctgggtgggtctggagtgcctaga 2939
Db 119010 GAGGAGAGAGACCCAGAGAGAGGGGAACGTGTGGGCCCTGGGTCTGGAGTGGCTTGA 119069
QY 2940 ggaagccagagcgccagagagagagatgggatgggtggcagcccagggctggggccgac 2999
Db 119070 GGAAGCCCAAGGCCAGAGAGAGAGATGGGATGGGTGGCGAGCCCCAGGCTGGGCCGAC 119129
QY 3000 ctcaactgtgctctgtgccccctgcctggaccaggtgagccatcgtgcocatcaagcc 3059
Db 119130 CTCACACTGTGCTGTGTGCCCTGCCCTGGACAGGTGAGCCATACGTCCGCATCAAGGC 119189
QY 3060 ctcaactgtgaggggggagcaggtgtccctgcctcaggggtgaagctgttgaagttaat 3119
Db 119190 CTACACTGCTGTGGAGGGGAGAGGTGCTCTGTCTGAGGGGTGAAGCTGTGTGAGTAAAT 119249
QY 3120 tcacaagctcctggcgcgctggtggtgtcatcaggtagaggggccctctccatccagagc 3179
Db 119250 TCACAAGCTCCTGGACGGCTGTGTGGTCTATCAGGTAGGAGGGCCCCCTCTCCATCCAGAGC 119309
QY 3180 acccatctgagtcaagcccccagccagcaggggtgtttagggatctggggtagctgtctcc 3239
Db 119310 ACCCATCTGTAGTCAGCCCCCAGCAGACGGGTGTGTAGGGATCTGGGTGATCTTGTCCC 119369
QY 3240 tgggactctggtagcaactgcccctctctgggcttagttccatctcagtagcagga 3299
Db 119370 TGGGACTCTGGGTAAAGCCACTGCCCTCTCTGGGCTTAGTTTCCATCTCAGTAGCAGGA 119429
QY 3300 gggatgagccccaccttgctgtcttggggatccaatgctctgttgcgaagtgggtgca 3359
Db 119430 GGGATGAGCCCCACCTTGCTCTGTGGGATCCAATGTCTTGTCCAAGTGGGTGCA 119489
QY 3360 tt 3361
Db 119490 TT 119491

RESULT 2

AC005098
LOCUS AC005098 230552 bp DNA PRI 07-OCT-2000
DEFINITION Homo sapiens BAC clone CTA-350L10 from 7q11.2, complete sequence.
ACCESSION AC005098
VERSION AC005098.2 GI:9211526
KEYWORDS HTG
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 230552)
Sulston,J.E. and Waterston,R.
Toward a complete human genome sequence
Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE 99063792
2 (bases 1 to 230552)
Theide,J., Abbott,A., Graves,T., Elliott,G. and Markovic,C.
The sequence of Homo sapiens BAC clone CTA-350L10
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 230552)
Waterston,R.H.

TITLE
JOURNAL
Submitted (12-JUN-1998) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE
AUTHORS
4 (bases 1 to 230552)
Waterston,R.H.
TITLE
JOURNAL
Direct Submission
Submitted (15-JUL-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE
AUTHORS
5 (bases 1 to 230552)
Waterston,R.
TITLE
JOURNAL
Direct Submission
Submitted (07-OCT-2000) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Jul 15, 2000 this sequence version replaced gi:3212893.
COMMENT

Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc>
Contact: sapiens@watson.wustl.edu

Summary Statistics

Center project name: H_RG350L10

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

The sequencing of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/CTB/CHR7>, send <mailto:egreen@nhgri.nih.gov>, or see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

Clone CTA-350L10 is from a release of the human BAC library CITB-HS-A. The library contains cloned DNA from human sperm. See: Shizuya et al., Proc. Natl. Acad. Sci. USA 89:8794-7 (1992); U-J. Kim et al., Genomics 34:213-8 (1996). The clone is available from Research Genetics, Inc. (<http://www.resgen.com>).

VECTOR: pBeloBAC11

Selection: chloramphenicol

NEIGHBORING SEQUENCE INFORMATION:

Actual start of this clone is at base position 1 of CTA-350L10;
actual end is at base position 230552 of CTA-350L10.

The sequence CTA-350L10 from base position 223330 to 222775 is a CA rich region. The sequence is not exact but it is believed to be the best representation of this region. The region was sized by PCR from clone DNA at 650 bp. The region corresponds to restriction digest hindiii: band size 7685 in silico and 7756 real.

FEATURES
source

1. 230552
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="7"
/map="7q11.2"
/clone="CTA-350L10"

repeat_region	/clone_lib="CITB-HS-A"
misc_feature	441..725 /rpt_family="Alu"
misc_feature	783..1149 /note="similar to EST BE000775 (NID:g8261008)"
misc_feature	910..1312 /note="similar to EST R22057 (NID:g776838) yh25f08.s1"
misc_feature	1286..1716 /note="similar to EST H72453 (NID:g1044269) ys04f07.r1"
repeat_region	1609..1864 /rpt_family="Alu"
misc_feature	1984..2271 /note="similar to EST R22113 (NID:g776894) yh25f08.r1"
misc_feature	2337..2682 /note="similar to EST AW383168 (NID:g6887827)"
misc_feature	2437..2558 /note="similar to EST N56845 (NID:g1200735) yw83e04.r1"
repeat_region	2532..2601 /rpt_family="MIR"
repeat_region	3054..3367 /rpt_family="Alu"
misc_feature	3240..3391 /note="similar to EST AI566941 (NID:g4525393) tn25b07.x1"
misc_feature	3588..3845 /note="similar to EST H70120 (NID:g1040326) ys01a09.r1"
misc_feature	3599..4017 /note="similar to EST H89808 (NID:g1080238) yu82g03.r1"
misc_feature	3741..3928 /note="similar to EST N56845 (NID:g1200735) yw83e04.r1"
misc_feature	3816..3932 /note="similar to EST AW523936 (NID:g7166321)"
misc_feature	3819..3932 /note="similar to EST AW503416 (NID:g7139965)"
misc_feature	3820..3932 /note="similar to EST AW197058 (NID:g6476288) xb15f09.x1"
misc_feature	3820..3932 /note="similar to EST AW163398 (NID:g6302431) au94g04.y1"
misc_feature	3871..3932 /note="similar to EST AI700706 (NID:g4988606) we39e12.x1"
misc_feature	3876..3932 /note="similar to EST T31784 (NID:g613882)"
misc_feature	3889..4387 /note="similar to EST T79157 (NID:g697666) yd70b07.s1"
misc_feature	4109..4387 /note="similar to EST AI696334 (NID:g4984234) tw59e05.x1"
repeat_region	4502..4812 /rpt_family="Alu"
repeat_region	5041..5087 /rpt_family="Mariner"
misc_feature	5254..5330 /note="similar to EST AA794004 (NID:g2856959) vr37a02.r1"
misc_feature	5265..5316 /note="similar to EST AW503416 (NID:g7139965)"
misc_feature	5265..5330 /note="similar to EST AI700706 (NID:g4988606) we39e12.x1"
misc_feature	5265..5330 /note="similar to EST AW197058 (NID:g6476288) xb15f09.x1"
misc_feature	5265..5330 /note="similar to EST AW523936 (NID:g7166321)"
misc_feature	5265..5330 /note="similar to EST T31784 (NID:g613882)"
misc_feature	5265..5348 /note="similar to EST AW163398 (NID:g6302431) au94g04.y1"
misc_feature	5284..5330 /note="similar to EST T30644 (NID:g612742)"
misc_feature	5305..5330 /note="similar to EST AW993840 (NID:g8254025)"
repeat_region	5750..6041 /rpt_family="Alu"
misc_feature	5964..6127 /note="similar to EST AA078486 (NID:g18379660)"
misc_feature	6032..6370 /note="similar to EST AI566941 (NID:g4525393) tn25b07.x1"

QY 1621 cttggggccctggcaggttgatgcccttggtctgacagggaaacacagaggaggagc 1680
|||||
Db 58830 CTTGGGGCCCTGGCAGGTTGTTGATGCCCTTGCTGGACAGGAAACACAGGAGGAGC 58889
QY 1681 agacagctcggagagtgaggagccagtggtctctgtgatatgtgagcaggttcagtg 1740
|||||
Db 58890 AGACAGCTGGGAGAGTGGGAGCCAGTGGTGTCTGTGATATGTGGCAGGTTCACTGG 58949
QY 1741 gaagctgaaggtatgacagaccttaggtccaggaaggaggtctgctgaaagtgggaca 1800
|||||
Db 58950 GAAGCTGAAGGATGACAGACCTTAGGCTCAGCAAGAGGGGTGCTGTGAAGTGGGGCA 59009
QY 1801 tcatcactgaccagaaaagggaacactggcagtgccagggctggatggggccctgcattgag 1860
|||||
Db 59010 TCATCACTGACCAGAAAGGGAACATGCGAGTGGCCAGGCTGGATGGGCGCTGCATTGAG 59069
QY 1861 cttgaaaaaactaataagaattggttaccattttattttattttattttattttattt 1920
|||||
Db 59070 CTTGAAAAAACTAATAGAAATTGGTTACCAATTTTATTTTATTTTATTTTATTTT 59129
QY 1921 tactttttgagatagagctcactcccttgctaaagctgagtgagtgctggttatctc 1980
|||||
Db 59130 TACTTTTTGAGATAGAGTCTCACTCCCTTGCTAAGGCTGGAGTGGGTGGTGTATCTC 59189
QY 1981 agtccactgcaacctctgctcccaggatcaagtgattctccagcctcagccctcccccagg 2040
|||||
Db 59190 AGCTCACTGCAACCTCTGCTCCCAAGATCAAGTCAATCTCCAGCCTCAGCCTCCCCAAG 59249
QY 2041 tagctgggattacaagcatgcaaccacatgctgctgataatttttttagttgaga 2100
|||||
Db 59250 TAGCTGGGATTACAAGCATGACCAACCATGCGCTGGATATTTTGTATTTTGTAGTTGAGA 59309
QY 2101 cgggtttcacaggttgccagactgctcgaacttctaactcaggtatctgactgctg 2160
|||||
Db 59310 CGGGTTTACCAAGTTGGCCAGACTGGTCTCGAACTTCTGACCTCAGGTGATCTGCCGTG 59369
QY 2161 cctggcctcccaagtgctggaattacagatgtgagccactgtccctggcctggttacc 2220
|||||
Db 59370 CCTCGGCCCTCCCAAGTGTCTGGAATTACAGATGTGAGCCACTGTCCTGGCGCTGTTACC 59429
QY 2221 cacattttaaaatgagtgatttacccttttatttgattttagatgagtttttttttt 2280
|||||
Db 59430 CACATTTTAAAATGAGTGATTTTACCCTTTTATGTGATTTTACAGCTGTGTTTTTTTTT 59489
QY 2281 ttttttgacaaaagtctggctctgtcccccaggtgagtgagtgagtaagtaactcagc 2340
|||||
Db 59490 TTTTGTGAGACAAAGTCTGGCTCTGTACCCAGGCTGGAGTGCAGTAAATGCAATCTCAGC 59549
QY 2341 tcactgcaaccttagctcctctgggtttcaagcaattctcctcagccactgagtgagc 2400
|||||
Db 59550 TCACCTGCAACCTTAGCCTCTCTGGGTTCAAGCAATTCTCTGCTCAGCCACTGAGTAG- 59608
QY 2401 ctggggttacaggcatgacacacacgacgagcaggttaatttttttttttttttttt 2460
|||||
Db 59609 CTGGGGTTTACAGGCATGACCAACACGCGCAGGCTAAATTTTGTATTTTGTAGTAGAGATG 59668
QY 2461 ggggttcgcatgttggccaggtggtctcgaactcctgactcaggtgagtcgcccacc 2520
|||||
Db 59669 GGGTTTCCCATGTGCGCCAGGCTGGTCTCGAACTTCCTGACTCAGGTGATCGCGCCCGC 59728
QY 2521 ttggcctcccaagtgctagattacaggttggaaccaccccttgccagcctgtggtctatc 2580
|||||
Db 59729 TTGGCCTCCCAAGTGTAGGATTACAGGTGGGAACCACTCGCCCGACGCTGTGCTATC 59788
QY 2581 gtttaaacactgggaagcctcgaagcccccagggccgacagttagctgagctgagcagtt 2640
|||||
Db 59789 GTTTAAACACTGGGAAGGCCTGACGCCCCCAGGGCGACAGTTAGTGTGAGCTGAGCAGTT 59848
QY 2641 cccagtgcaagtagagatgctccaccacactactatgctgagctatctgtgcatagt 2700
|||||
Db 59849 CCCAGTGCCAGGTACCGATGCTCCAGCCACCTACTCATGCTGATCTCTGTGCATAGT 59908

QY 2701 gaagtgtgggacagaccttcatcttattgggatctctggtccccagagtggttgccaat 2760
|||||
Db 59909 GAAGTGTCTGGACAGACCTTTCATCTTATGATCTCTGGTCCCCCAGAGTGGGTGGCAAT 59968
QY 2761 gaatgggaagtgaacaagctcactctgggtgtatgggggacagagggccgaagtctcagagtga 2820
|||||
Db 59969 GAATGGGATGAGACAAAGCTCACCTGGGTGTAGGGGCGACAGGGCCCAAGTCCACAGTGTGA 60028
QY 2821 cccccagagtggttccacagcagagcttgcacagagatctggatggacacagaggtgg 2880
|||||
Db 60029 CCCCCAGAGTGGTCCACAGAGAGCTTGCCAGAGGATCTGGGATGGACAGAGGGTGG 60088
QY 2881 agggaggagaccccaagagggggaactgtggccctgggtggtctgagtgctggag 2940
|||||
Db 60089 AGGGAGGAGACCCAGAGAGGGGGAACCTGTGGCCCTGGGTGGGTCTGGAGTGCCTGGAG 60148
QY 2941 gaagcccaggccagagaggaggaagatgggagtgggtggcagccccaggtctgggcccagcc 3000
|||||
Db 60149 GAAGCCCAAGGCCAGAGAGAGATGGGATGGGTGGCGAGCCCGAGCTGGCGCGACC 60208
QY 3001 tcacactgtctctgtgccccctgcccgtggaccaggtgagccatcagtcgccaatcagacc 3060
|||||
Db 60209 TCACACTGTGCTCTGTGCCCTGCCCCGTGGACAGGTGAGCCATACGTGCCCATCAAGGCC 60268
QY 3061 tacactgtgtgaggggagacaggtgtccctgctcaggggtgaagctgttgagtgaaatt 3120
|||||
Db 60269 TACACTGTGTGGAGGGGACGAGGTGTCCTGCTCGAGGTTGAAGCTGTTGAGGTCTATT 60328
QY 3121 cacaagctcctggacggcttgggtgcatcaggttagggagggccccctctccatccagagaca 3180
|||||
Db 60329 CACAAGCTCTGGACGGTGGTGGTTCATCAGGTAGGAGGGCCCCCTCTCCATCCAGAGCA 60388
QY 3181 cccatctcagtcagccccagccagcagcaggggtgttttagggatctgggtgactgtacct 3240
|||||
Db 60389 CCCATCTGAGTCAGCCCCAGCAGGACGGGGTGTATTAGGGATCTGGGGTGAATTTGCCCT 60448
QY 3241 gggactctgggttaagccactgccccctctctgggttagtttccatctcagtagcagggag 3300
|||||
Db 60449 GGGACTCTGGGTAAAGCCACTGCCCTCTCTGGGCTTAGTTTCCATCTCAGTAGCAGGAG 60508
QY 3301 ggatgagccaccttgctgctcttgggtggggtacaaagtcccttgcacagtgaggtgcat 3360
|||||
Db 60509 CAATGAGGCCACCCCTGCTGTCTTGTGGGATCCAATGTCTCTTCCAAAGTGGGTGCAT 60568
QY 3361 t 3361
|
Db 60569 T 60569

RESULT 3
AC004883
LOCUS AC004883 131359 bp DNA PRI 21-DEC-1999
DEFINITION Homo sapiens PAC clone RP4-771P4 from 7q11.21-q11.23, complete sequence.
ACCESSION AC004883
VERSION AC004883.2 GI:4263746
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 131359)
AUTHORS Sulston, J. E. and Waterston, R.
TITLE Toward a complete human genome sequence
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE 99063792
REFERENCE 2 (bases 1 to 131359)
AUTHORS Kalicki, J. and Laplant, Y.
TITLE The sequence of Homo sapiens PAC clone RP4-771P4
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 131359)
AUTHORS Waterston, R. H.
TITLE Direct Submission


```
73330. .73513.74993. .75021.76304. .76345.77444. .77485,
78262. .78320)
/ gene="GTF21"
/ note="match to AF035737 (PID:g2827180); H_DJ0771P04.1"
/ codon_start=1
/ product="general transcription factor 21"
/ protein_id="AAD15419.1"
/ db_xref="gi:4263747"
/ translation="MAOVANSTLPEDEEESRMVTVFLMSALESCKELAKSKAEV
ACIAVYEDVFCVGTGRFVNTRKDFKDFKVCVEEKEAEKHKHSTTQANRM
SVDAVEIETLRKTVEDYFCYCYKALGKSTVVPYKELRLDQSAVVVQGLPEGVAK
HPNVDLALIKWILENAAGTSFIIKRFPLEKRWKVGQRMVMDADRSLSPGSKGPI
KVTPEPDSGISLEMAAVTVKEESDPDIYKNIQAPSETDDVDLQPLKSLGCS
HSESENEGTEMEYPAEDSTQHVSETSEDEVEVETIEDDDYSPSKRPKANLELPQP
VPEPANAGKRKRVENFEKNARITDLRKQVEELFERKYAQAIKAKGVTIPYPLFOS
HEDLYVVEGLPEGIPFRPSTYIPRLERILLAKERIRFVKKHELLNSTREDILODK
PASGVKEBWARIITKLKRWDDLFCKFAEALGSTEAKAPYQKFEAPHPNDLYVEGLP
ENIFRPSWTFGIPLREKLIQVGNRIKFIKRPBELLTHSTVEYTPQPRINTPVKEDWN
RITLKRQVEEIFNLKFAQALGLTEAVKVPYVPFESNPEFLYVEGLPEGIPFRPTWP
GIPRLERIVRGSNIKFKVVKPELVIYLPFGMAKINTKALQSPKRPSPGSKVP
EIEVTEGPNNNPOTSATVPTOTNGSNVPEKPRGREFSFEAWNAKITDLKQVENL
FNEKCGALGLQAVKVPFALFESFPEDFVVEGLPEGVPRRPSTFGIPREKILRNK
AKIRFLIKPMPETAIKESTSSKSPRKINSSPNVNTASGVEDLNIQVTFIPDDDN
ERLSKVRAROLRQVNDLFSRKEAGEAIGMGPVVKYPIKITINPNCVQVDMGPPGV
FKAPSYLEISMRRLDSAEFIKFTVIRPFPLVNNQLVDQSESGPVIQESAEPQ
LEVPAEIEIKETDSSQIKQEPDPTW"
misc_feature
/ gene="GTF21"
/ note="match to EST AA490549 (NID:g2219722) aa42e07.rl"
8640. .8824
/ gene="GTF21"
/ note="similar to EST AA609881 (NID:g2458309) af08d04.s1"

Query Match          98.3%; Score 3304.2; DB 9; Length 131359;
Best Local Similarity 99.3%; Pred. No. 0;
Matches 3339; Conservative 0; Mismatches 18; Indels 4; Gaps 2;

Qy      1  aaaaagtcaggcactgtgctcatcctgtaataccagcaccttggggagccgagcggt 60
Db 101664  AAAAGGTCAGGCACACTGTGCCTCATCGCTGTATCCAGCACTTTGGAGGCGGAGCGGT 101723

Qy      61  tgaatacactgaggtcaggagtcagaccagcctgaccaatatcgtgaatacctcatc 120
Db 101724  TGAATACACTGAGGTCAGGAGTTCGAGACCAGCAGCTGACCAATATCCTGAAATCCATCTC 101783

Qy      121  tactaaaaatacaaaattagccaagtgtgtggcggtgtgttgatcccaagctacttg 180
Db 101784  TACTAAAAATACAAAAATTAGCCAGGTGTGTGGCGGGTGTCTGTAGTCCCAAGCTACTTG 101843

Qy      181  ggaggtgagcgagtgtaattacttgaacctgggaggtggaggttgcaatgagccaaagt 240
Db 101844  GGAGGCTGAGGCAGGTGAATTACTTTGAACCTGGGAGGTGGAGGTGCAATGAGCCAAAGAT 101903

Qy      241  tgcaccactgcactccagtgcagagcgagactccatctcaaaaaaataaaaaaaag 300
Db 101904  TGCACCACTGCACCTCCAGTGACAGCGAGACTCCATCTC---AAAAAAAAAAAAAAAAAG 101960

Qy      301  ttgggaaaagccaggtgcagtggtccacgcctgtaatacccaacactttaagaggtga 360
Db 101961  TTGGGAAAAGGCCAGGTGCAGTGGCTCCACGCTGTAAATCCCAACACTTTAAGAGGCTGA 102020

Qy      361  ggtggagaatcctttgagccaggagttcagagaccagcctgggcattgtgtccaaagcct 420
Db 102021  GGTGGAGAAATCCTTTGAGCCAGGAGGTTCGAGACCAGCCTGGGGCATTTGCCAAGACCT 102080

Qy      421  tgtctttcaaaaaataatagccgggtgtgtggcatacgtctctgtgtccacagctatccgg 480
Db 102081  TGTCTTTACAAAAAATTAGCCGGGTGTGTGTCATACGTCATATGTGTCCACGACTATTCGG 102140

Qy      481  aggctgagggcagggagatgtcttgagccttagagtcctaggtggtctgtagctgtgata 540
Db 102141  AGGCTGAGGCAGGGAGATTGCTTGAGCCTAGGAGTCCAGGGCTGTAGTGAGCTGTGATCA 102200
```

```
Qy      541  cgtcactgtactctagcctggggaacacagcagcaagactctgtctccaaaaaataaaa 600
Db 102201  CGTCACTGTACTCTAGCCTGGGCAACAGAGCAAGACTCTGTCTCCAAAAAAGANAATAA 102260

Qy      601  gttggaaaagctcactaacttcatcagatgagaacaagacatgttgaagtgtaggc 660
Db 102261  GTTGGAAAGGCTCACTAACTTCATCAGATGAGAACAAAGGACATGTTTGAAGTGTGAGC 102320

Qy      661  cgaacccctggagaacgctatgcgccagagaaatcagggcagcagacagatcaagatgca 720
Db 102321  CGAACCTGGAGAACGCTATGCGCCAGAAATGAGGGCAGCAGAGACTCAAGATGCCA 102380

Qy      721  ggcctgttctggagccccagatgggcccctgcaatgcccactcaccctgcctccctctt 780
Db 102381  GGCCTGTCTTCTGGAGGCCAGATGGCCCTGCAATGCCACTCACCTTGCCCTCCTCTT 102440

Qy      781  gccccagacataccacggcccccatactctctgcagacgtaccgcgcatctgcgactacgag 840
Db 102441  GCCCCAGACATCACCGGCCCATCATCTCTGCAGAGCTACCGGCCCATTTGCCAACTACGAG 102500

Qy      841  aagacctcggactccgaatagctctgtccacggggagcgtgtgtaggtcgtgagaag 900
Db 102501  AAGACTCGGGCTCCGAGATGGCTCTGTCCACGGGGAGCTGGTGGAGTCTGTAGAGAG 102560

Qy      901  agcgaagcgggtcagacacctcccacacctacggggctccttcccctgtgtgctcaggaacca 960
Db 102561  AGCGAGAGCGGTACAGACCTCCACCTTACGGGGCTCTTCCCTGGTCTCAGGAACCCA 102620

Qy      961  cagccaaagcccccgtgcaaggctcagcagcctgagcctggcccctgggagactccagctctg 1020
Db 102621  CAGCCACAAGCCCCCTGGCAAGGCTCAGCGACGCTTGGCCCTGGGAGGACTCCGCTCTG 102680

Qy      1021  ttaggggccctaaatgtctctccacacctgtgggtcgctctctctcttagtgtcacct 1080
Db 102681  TTAGGGGCCCTAAATGTCTCTCCACACACTGTGGGTGCGCTTCTGTCTTAGTGTGACCCCT 102740

Qy      1081  gtggtggctgtgggcactctgtgcatggcagggccggggcgggcagtcgtctgctgtctgt 1140
Db 102741  GTGGTGGCTGTGGGCATCTGTGCATGGCAGGCGCGGGCGGGGCATGTCTGCGTGTCTGT 102800

Qy      1141  ctggatgggtatgggaacgctctgttcatatgaagtgggctcagagcgtgtgattctgtga 1200
Db 102801  CTGGATGGGTATGGGACCGCTCTGTTTATTGAAGTGGGCTCAGAGCTGTGATTCTGTGA 102860

Qy      1201  gcattgtgcacatgcacatgtgacctcattgtccagttggtgaggtgacatttccaa 1260
Db 102861  GCATTGTGCATGCATGCATGTGACCTCATTTGTCAGTGTGGTGAAGGTGACATTTCCAA 102920

Qy      1261  atctgagcattggacatcagttgtctgtgtccctgtgtccctcaccatccctgtatggctg 1320
Db 102921  ATCTGAGCATTTGGACATCAGTGTGTCTGTGTCCCTGTGTCTCTCACCATCCCTGATGGCTG 102980

Qy      1321  caggagccgctgggcccctgcccctcagtcacattcccgcacctctggcacaggttggtg 1380
Db 102981  CAGGAGCCGCTGGGCCCTGCCCTCAGTCACTATCCCGCACCTCTGGCACAGTGTGGTG 103040

Qy      1381  gttctgtcagatgaaagcaaaagcagagctggtatcccgacatccttccctcgagccctgga 1440
Db 103041  GTTCTGTCAAGTGAAGCAAAAGCAGGCTGGATCCACGCGTCTTCTCGAGCCCTCGA 103100

Qy      1441  cagtcctgacgagcagcagaccctgagcccaactatgcaggtgccccctgcctcccgag 1500
Db 103101  CAGTCCTGACGAGCAGGAGACCCTGAGCCCAACTATGAGAGTGGCCCCCTGCCCTCCGAG 103160

Qy      1501  gctgtagggggtgtgggaaaaagggcagcaggggctcagggaatttgtagtgcactcttg 1560
Db 103161  GCTGTAGGGGTGTGGAGAAAGGGGACGAGGCTCAGGAGATTTGAGTGACTGCTTTG 103220

Qy      1561  gagctcgggctggttgccttggccttggcagaaaaagtcagggcctaagatctcatcggtctg 1620
Db 103221  GAGCTGGGCTGTTGCTGGCTTGGCAGAAAAAGTCAGGGCTAAGATCTCATCGACTCTGG 103280

Qy      1621  cttggggggccctggcaggttgtgtatgacctgtgtgtgagcaggggaaccagggaggagc 1680
```

|||||
Db 103281 CTTGGGGCCCTGGCAGGTTGTGATGCCCTTGGCTGGACAGGGAACACGAGGAGGAGC 103340
Qy 1681 agacagactcggagagtgaggagccagtggtctctgttgatattgtgcccaggtcagtg 1740
|||||
Db 103341 AGACACTGGGGAGAGTGGGAGCCAGTGGTGTCTGTGTGATATGTGGCAGGTTTCAGTGG 103400
Qy 1741 gaagctgaaggtatgagcagaccttagctcaagaaaggggctgcctggaagtggggca 1800
Db 103401 GAAGCTGAGGATGAGCAGACCTTAGCTCAGGAAGGAGGGCTGCCTGGAAGTGGGGCA 103460
Qy 1801 tcactcagtcacagaaagggaaactcggcagtcagggctggtatggggccctgcattgag 1860
Db 103461 TCATCACTGACCAGAAAGGAAACTGCGAGTGCAGGGCTGATGGGGCCCTGCATTGAG 103520
Qy 1861 cttgaaaaaactaataagaattggttacaattttattattattattattattattatt 1920
Db 103521 CTTGAAAAAATAATAGAAATGGTTACCAATTTATTTATTTATTTATTTATTTATTT 103580
Qy 1921 tactttttgagatagagctcaectcccttgctaaggctggagtcggtggtgctatctc 1980
Db 103581 TACTTTTTGAGATAGATCTCACTCCCTTGCTAAGTTGGAGTGGGTGCTATCTC 103640
Qy 1981 agtccactgcaacctctgctccaggatcaagtgatctctccagcctcagcctcccccagg 2040
Db 103641 AGTCACTGCAACCTCTGCCTCCAGGATCAAGTGATTCTCCAGCCTCAGCCTCCCAAG 103700
Qy 2041 tagctgggattacaagcatgacacacatgctggatgataattttttatttttagttgaga 2100
Db 103701 TAGCTGGGATTACAAGCATGCAACCACTGCGCTGGATATTTTGTATTTTAGTTGAGA 103760
Qy 2101 cgggggttcaacaggttgccagacgtgctctgaactctctgaactcagtgatctgcctg 2160
Db 103761 CGGGGTTTACCAGGTTGGCCAGACTGCTCGAATCTTCTGACCTCAGTGATCTGCCTG 103820
Qy 2161 cctcggcctcccaagtgctggaaattacagatgtagcgaactgtccctggcctggttacc 2220
Db 103821 CCTCGGCCTCCCAAGTGTGTGAATACAGATGTGAGCACTGTCCCTGGCGCTGTACC 103880
Qy 2221 cacattttaaatgagtgattctaccccttttatgtgatttaccagctgttttttttt 2280
Db 103881 CACATTTTAAATGAGTGATTTTACCCCTTTTATGTGATTTTACAGCTGTGTTTTTTT 103940
Qy 2281 ttttttgacaaaagtctggctctgtcaccagggctggagtgagtaaatgcaatctcagc 2340
Db 103941 TTTTTTGACAAAACCTCTGGCTCTGTCAACCAGGCTGGAGTGCAGTAATGCAATCTCAGC 104000
Qy 2341 tcactgcaaccttagcctcctgggttcaagcaattctcctcagcccaactgagtagc 2400
Db 104001 TCACCTGCAACCTTAGCCCTCTGGGTTCAAGCAANTCTCTGCCTCAGGCACCTGAGTAG- 104059
Qy 2401 ctgggggttacaggcatgaccaccacgcagagctaatttttgtatttttagtagagatg 2460
Db 104060 CTGGGGTTACAGCATGACACCAACGACGCGAGCTAATTTTGTATTTTGTATGATG 104119
Qy 2461 ggggttcgcatgttggccaggtggtctgaactccctgacctcaggtgagtcgcccgc 2520
Db 104120 GGGTTTGGCCATGTTGGCCAGGCTGCTCGAATCTCTGACCTCAGGTGATCCGCCCGCG 104179
Qy 2521 ttggcctcccaagtctaggtattacaggtgggaacaccccttgcccagccttggtctatc 2580
Db 104180 TTGGGCTCCCAAGTGTCTAGGATTAAGTGGGAACCACTCGCCAGCCTGTGGCTATC 104239
Qy 2581 gtttaaacactgggaggtcagccccccagggccgacagttagctcagctgagcagtt 2640
Db 104240 GTTTAAACACTGGGAAGGCTGCACGCCCCAGCCGACAGTTAGTGTGACGTGACAGTT 104299
Qy 2641 cccagtgcaggtagacgagtagctccacccactactcagctgagctatctttgttcagtt 2700
Db 104300 CCCAGTGCAGGTAGACGGATGCTCCAGCCACCTACTCATGCTCATCTCTGTGTATAGT 104359
Qy 2701 gaagtgtctgacacaccttcatcttatggtatctctgttccccagagtggttggaat 2760
|||||

Db 104360 GAAAGTGTCTGGACAGACCTTCATCGTTATTGGGATCTCTGGTCCCCAGAGTGGTGGCAAT 104419
Qy 2761 gaattggagttggaacagctcaacctgggtgttagggggcagagggccgaagtccagaagtata 2820
|||||
Db 104420 GAATGGGAGTGGACAAGCTCACCTGGGTGTAGGGGGCAGAGGGCCGAAGTCCAGAGTGT 104479
Qy 2821 cccccagagtggtgtccagcagagagcttgcgcagagagatctctggatggagcagaggggtg 2880
Db 104480 CCCCAGAGTGGTGGTCCAGCAGGAGCTTCCAGAGGATCTGGATGGAGCAGGAGGGTGG 104539
Qy 2881 agggaggagaccagaagagggggaactgttggccctgggttgggtctggaagtgcctggag 2940
Db 104540 AGGGAGGAGAGCCACAGAGGGGGAACCTGTGGCCCTGGTGGTGGTGGTGGTGGTGGAG 104599
Qy 2941 gaagcccaaggcagagagagaagatggatgggtgagcagccccagagctggggccgacc 3000
Db 104600 GAAGCCCAAGGCGCAGAGAGGAGAGATGGATGGGTGGGAGCCCCAGGCTGGGCCGACC 104659
Qy 3001 tcacactgtctctgtgccccctgccgtggacaggtgagccatcagtcgcatcaaggcc 3060
Db 104660 TCACACTGTGCTGTGCTGCTGCCCTGCCGTGACAGGTGAGCCATAGTGCCTCATCAAGGCC 104719
Qy 3061 tacactgtctggaggggggacgaggtgtccctgtcgtcaggggtggaagctgttggagtaatt 3120
Db 104720 TACACTGTGTGGAGGGGAGCAGGTGTCCCTGTCTCGAGGGTGAAGCTGTTCAGGTCAAT 104779
Qy 3121 cacaactccttgagcgcgtggtgggtcatcaggttagagggccctctccatccagagca 3180
Db 104780 CACAAGCTTCTGGACGGCTGGTGGGTTCATCAGGTAGGAGGGCCCTCTCCATCCAGAGCA 104839
Qy 3181 cccatctgagtcagccccagcagcagcgggtgtttagggatctctgggtgagctgtccct 3240
Db 104840 CCCATCTCAGTCAGCCAGCCAGCAGGCGGGTGTTCAGGATCTGGGTGACTTGTCCCT 104899
Qy 3241 gggactctgggttaagccactgccctctcctgtcgtcaggttagttccatcctcagtagcaggag 3300
Db 104900 GGGACTCTGGGTAAAGCCACTGCCCTCTCTGGGCTTAGTTTCCATCTCAGTAGCAGGGAG 104959
Qy 3301 ggaagccccacctgctctcttgggaatccaatgctccttgcctcgaagtggtgcat 3360
Db 104960 GAATGAGCCCAACCTTCCTGCTGTGGGATCCAAATGTCCTTGTCCAGTGGGTGAT 105019
Qy 3361 t 3361
Db 105020 T 105020

RESULT 4

AC083884
LOCUS AC083884 169604 bp DNA PRI 23-MAY-2001
DEFINITION Homo sapiens clone RP11-813J7, complete sequence.
ACCESSION AC083884
VERSION AC083884.6 GI:14190780
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 169604)
AUTHORS Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE
2 (bases 1 to 169604)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (04-OCT-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE
3 (bases 1 to 169604)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (23-MAY-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,

MO 63108, USA
On May 23, 2001 this sequence version replaced gi:13431264.
Center project name: H_NH0813J07.
FEATURES
Location/Qualifiers
1. 169604
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RP11-813J7"
BASE COUNT 44465 a 38383 c 38794 g 47962 t
ORIGIN

Query Match 98.2%; Score 3299.4; DB 9; Length 169604;
Best Local Similarity 99.3%; Pred. No. 0;
Matches 3336; Conservative 0; Mismatches 21; Indels 4; Gaps 2;

Qy 1 aaagctcaggcactgtgctcattccctgtatccacagcactttggaggccgagcgt 60
|||||
Db 104774 AAAGGTCAGGCACCTGTGGCTCATCGCTGTATATCCAGCACCTTTGGAGGCGGAGCGGT 104833
|||||
Qy 61 tgaatcacctgaggtcaggagttcgagaccagcctgaccaatattcgtgaaacctccatctc 120
|||||
Db 104834 TGAATCACCTGAGGTCAGGAGTTCGAGACCAGCCTGACCAATATCGTGAATCCATCTC 104893
|||||
Qy 121 tactaaaaatacaaaaattagccaggtgtgtggtggtgtgtgttagtcccgactcttg 180
|||||
Db 104894 TACTAAAAATACAAAATTAGCCAGGTGTGTGGCGGGTGTGTAGTCCCGAGCTACTTG 104953
|||||
Qy 181 gtagctgagcagtgtaattacttgaacctggagggtgaggttgcaatgagccaaagt 240
|||||
Db 104954 GGAGGCTGAGGCAGGTGAATTACTTGAACCTGGGAGGTGGAGGTTGCCAATGAGCCAAAGT 105013
|||||
Qy 241 tgcacctgcaactccagtgcagagcgagactccatctcaaaaaaataaaaaaaag 300
|||||
Db 105014 TGCACCACTGCACTCCAGTGACAGAGCAGACTCCATCTC --- AAAAAAATAAAAAAAG 105070
|||||
Qy 301 ttgggaaagccaggtgcagtggtgtccacgctgtatcccaacactttaagagctga 360
|||||
Db 105071 TTGGGAAAAGGCCAGGTGCAGTGGCTCCAGGCTGTAATCCCAACACTTTAAGAGGCTGA 105130
|||||
Qy 361 gttggagaactctttgagccaggagttcgagaccagccttgaggcttggtcccaagaact 420
|||||
Db 105131 GGTGGGAATACTTTGAGCCAGGAGTTCGAGACCAAGCCTGGGCATTGTCCCAAGACT 105190
|||||
Qy 421 ttctttcaaaaaattagccgggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 480
|||||
Db 105191 TGTCTTTACAAAAAATAGCCGGTGTGGTGCAATACGTCTATGTGTCCAGCTATTCGGG 105250
|||||
Qy 481 aggtgagggcaggagattgttgagcctaggagttcagggtgtagtgtgtgtgtgtgtgtgt 540
|||||
Db 105251 AGGCTGAGGCAGGAGATTGCTTGAGCCTAGGAGTCCAGGCGTGTAGTGAGCTGTATCA 105310
|||||
Qy 541 cgtcaactgtactctagcctgggcaacagagaagactctgtctccaaaaaagaaataaaa 600
|||||
Db 105311 CGTCACTGTACTTGTAGCTGGGCAACAGAGCAAGACTCTGTCTCCAAAAAAGAAATAAAA 105370
|||||
Qy 601 gttggaaaggtcactaactcatcagatgagacaagaacatgtttgaagttgagac 660
|||||
Db 105371 GTTGGAAAGGCTCACTAATTCATCAGATGAGACAGAGACATTTTCAAGTGTGAGGC 105430
|||||
Qy 661 cgaagcctggaaacgctatgcgccagagaaatgcaggccagcagcagactcaagatgcca 720
|||||
Db 105431 CGAAGCCTGGAGAACGCTATGCGCCAGGAAATGCAGGCGACGACAGACTCAAGATGCCA 105490
|||||
Qy 721 ggcctgttctggagggccagatgggcccctgcaatgcccactcaccctgccctcccttt 780
|||||
Db 105491 GCGCCTGTCTGGAGGCCAGATGGGCCCTGCAATGCCCACCTACACCTGCCCTCCCTCTT 105550
|||||
Qy 781 gcccagacatcacggcccactatctcagacgtaccggtccattgccagctacag 840
|||||
Db 105551 GCCCCAGACATACCCGGCCCCCATCTCTGAGAGCTACCGGCCATTTTGCCAACTACAG 105610
|||||
Qy 841 aagacctcggctcagatggtctctgccacgggggagcgtggtgaggttcgtgagag 900
|||||

Db 105611 AAGACTCGGGCTCGAGATGGCTCTGTCCACGGGGAGCGTGGTAGGTCGTAGAGAAG 105670
|||||
Qy 901 aqcgagagcgttcagacacctccacacttacggggtcccttcccctggtgtcctcaggaacca 960
|||||
Db 105671 AGCGAGGGGTGACAGCTCCACCTTACGGGGCTCCCTTCCCTGTGTCTAGTAGTGCACCT 105730
|||||
Qy 961 cagccacaagccccctgccaaaggtcagagcctggccccctgggagggactccagctctg 1020
|||||
Db 105731 CAGCCACAAGCCCCCTGCCAAGGCTCAGCAGCCTTGCCCCCTGGGAGGACTCCGGCTCTG 105790
|||||
Qy 1021 ttaggggccctaaatgtctctcccaactgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1080
|||||
Db 105791 TTAGGGGCCCTAAATGTCTCCACACTGTGGGTGGCGCTTCTGTCTTGTAGTGCACCT 105850
|||||
Qy 1081 gtggtggt 1140
|||||
Db 105851 GTGGTGGCTGTGGGCATCTGTCTTGGCAGGCGGGGGGGGCATGTCTGCGGTGTCTGT 105910
|||||
Qy 1141 ctggatggt 1200
|||||
Db 105911 CTGGATGGGTATGGGACCGCTCTGTTCATTTATGAAGTGGGCTCAGAGCTGTGATTCTGTA 105970
|||||
Qy 1201 gcatgtgtcatgtcatgtcatgt 1260
|||||
Db 105971 GCATGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 106030
|||||
Qy 1261 atctgagcatgtgacatgt 1320
|||||
Db 106031 ATCTGAGCATTTGGACATCAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 106090
|||||
Qy 1321 caggagcgctgt 1380
|||||
Db 106091 CAGGAGCGCTGTGGCCCTGCCCTCAGTCACATTTCCCGCACCTCTGGCACAGTTGGTG 106150
|||||
Qy 1381 gttctgtcagatgaagcaagcaggt 1440
|||||
Db 106151 GTTCTGTGATGAAAGCAAGAGGAGGTGGATCCAGGCTCTTCTCTCGAGCCCTGGA 106210
|||||
Qy 1441 cagtcctgacagagcaggaagacctgagcccaactatgcagggtgccccctgccccctccgag 1500
|||||
Db 106211 CAGTCTCAGCAGAGCGGAAGACCCCTGAGCCCAACTATGCAGGTGCCCTTGCCCTCCGAG 106270
|||||
Qy 1501 gctgtagggt 1560
|||||
Db 106271 CCTGTAGGGGTGTGGGAAAGGGGCGAGCGGCTCAGGGATATTGAGTGAATGCTTTG 106330
|||||
Qy 1561 gagt 1620
|||||
Db 106331 GAGTCTGGCTGT 106390
|||||
Qy 1621 ctgtggggccctgtgaggt 1680
|||||
Db 106391 CTGTGGGGCCCTTGGCAGGTTGT 106450
|||||
Qy 1681 agacactcggagagtggt 1740
|||||
Db 106451 AGACACTTGGGAGAGTGGGAGGCCAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 106510
|||||
Qy 1741 gaagctgaagatgagcagacacttaggttcaggaagggggtgtgtgtgtgtgtgtgtgtgtgt 1800
|||||
Db 106511 GAAAGTGAAGGATGAGCAGACCTTAGGCTCAGGAAGGAGGCTGCCTGGAAGTGGGGCA 106570
|||||
Qy 1801 tcatcactgaccagaaagggaaaaactggcagtgccagggctgtgtgtgtgtgtgtgtgtgtgt 1860
|||||
Db 106571 TCATCACTGACCAGAAAGGAAACCTGGCAGTGCAGGCTGGATGGGCGCTGCAATTGAG 106630
|||||
Qy 1861 ctgtgaaaaactaataagattgttaccattttattttattttattttattttattttatttt 1920
|||||
Db 106631 CTTGAAAAAACTAATAAGAAATGTTACCATTATTTATTTATTTATTTATTTATTTATTTAT 106690
|||||
Qy 1921 tacttttttgatagatgactcactcccttgcctaaggtcggaggtggtgtgtgtgtgtgtgtgt 1980
|||||

Db 106691 TACTTTTTTGTAGATAGAGTCTCACTCCCTTGCTAAGGTTGGAGTGGGTGCTATCTC 106750
Qy 1981 acctcaactgcaactctcctccccaagatcaagtgaattctccagcctcagcctcccagg 2040
Db 106751 AGCTCAGTCAACCTCTGCCTCCCAAGGATCAAGTGAATTCCTCAGCCTCAGCCTCCCAAG 106810
Qy 2041 tagctgggattacaagcatgcaaccaccatgctggtgataatttttttagttgaga 2100
Db 106811 TAGCTGGGATTACAGCATGCACCAACCATGCTGGATTAATTTTGTATTTTAGTTGAGA 106870
Qy 2101 cggggtttcccaagattggccagactggtctcogaactctgacctcaggtgatcgcctg 2160
Db 106871 CCGGGTTTACCAGGTTGCCAGACTGGTCTCGAACTTCTGACCTCAGGTGATCTGCCTG 106930
Qy 2161 cctcgacctcccaagactggaattacagatgtagccactgtccctggcctgattacc 2220
Db 106931 CCTCGGCTTCCCAAGTGTGGAATACAGATGTAGGCACACTGTCCCTGGCGTGTATACC 106990
Qy 2221 cacattttaaaatggagtgaattccacctttttatggtattacagctgtttttttttt 2280
Db 106991 CACATTTTAAATGAGTGATTTTACCCCTTTTATCTGGATTACAGCTTGTTTTTTTTT 107050
Qy 2281 ttttttgacaaaagtctggtctctgcacccaggtctgcacccaggtcagtgacgaatctcagc 2340
Db 107051 TTTTGTGACAAAAGTCTGGCTCTGTCAACAGGCTGGAGTGCAGTAATGCAATCTCAGC 107110
Qy 2341 tcactgcaaccttagcctcctggttcaagcaattctcctcctcagccacctgagtagc 2400
Db 107111 TCAGTGCAACCTTAGCCTCTGGGTTCAGCAATTCCTTGCTCAGCCACCTGAGTAG- 107169
Qy 2401 ctggggttacaggcatgaccaccaccgcccaggtctaattttttgtatttttagtagagtg 2460
Db 107170 CTGGGTTTACAGCATGCACCAACCAACGCGCAGGCTAATTTTGTATTTTAGTACAGATG 107229
Qy 2461 ggggttcgcattgttggccaggtcgtctcgaactcctcgaactcaggtcaggtcgcgcgc 2520
Db 107230 GGGTTTGGCATATTTGGCCAGAGCTGTCTCGAACTTCCTGACCTCAGGTGATCGGCCCGG 107289
Qy 2521 ttggcctcccaagtgcctaggtacaggtgggaaccacctggccagcctgtgctatc 2580
Db 107290 TTGGCCTCCCAAGTGTAGGATTCAGGTGGGAACCACTCTGCCACGCTGTGGCTATC 107349
Qy 2581 gtttaaacactgggaagcctcagccccccagccagtcagttagctcagctgagcagtt 2640
Db 107350 GTTTAAACACTGGGAAGGCTCAGCCCCCAGCGCCGACAGTTAGTGCAGCTGACAGTT 107409
Qy 2641 cccagtcaggttagcaggtatctcaccaccactactcaggtgatctcttctgtcatagt 2700
Db 107410 CCCAGTGCAGGTAGACGGATGCTCCAGCCACTACTCATGGCTGATCTTTGTCTATAGT 107469
Qy 2701 gaagtgtctggacagaccttcattgtatgggatctctggtcccccagagtggtgccaat 2760
Db 107470 GAAGTGTCTGGACAGACCTTCATCGTTATGGGATCTCTGGTCCCCAGAGTGGGTGGCAAT 107529
Qy 2761 gaatggagtgacaaagctcacctgggtgtaggggagagggccgaagtcocagagtgta 2820
Db 107530 GAATGGGAGTGACAAAGCTCACCTGGGTAGGGGAGAGGGCCGAACTCCAGAGTGTA 107589
Qy 2821 cccccagagtggttgcacagcagagcttgcagggagatcttggatggagcaggaaggttgg 2880
Db 107590 CCCCCAGAGTGGGTGCCACAGAGAGCTTGCCAGGGGATCTGGGATGGAGCAGAGGGGTGG 107649
Qy 2881 agggagagaccagaagaggggaactgtggccctgggtgggtctgagtgacctgag 2940
Db 107650 AGGGAGGAGACCCAAAGAGGGGAACGTGTGGGCCCTGGGTGGGTGTGGAGTGCCTGGAG 107709
Qy 2941 gaagccaggtcagagagaggaagatgggatgggtggcagccccaggtgggcccagcc 3000
Db 107710 GAAGGCCAGGCCAGAGAGAGAGATGGGATGGGTGGCGAGCCCCAGGCTGGGCCGACC 107769
Qy 3001 tcacactgtcctgtgcccctgcctggagccaggtgagccatcagtcacatcaagccc 3060
Db 107770 TCACACTGTGCTCTGTGCCCTTGCCTGGACAGGTGAGCCATAGTGCCTCATCAAGGCC 107829

Qy 3061 tacactgtctggaggggacgaggtgtccctgctcagaggtgaagctgttgaggttaatt 3120
Db 107830 TACACTGCTGTGGAGGGGACGAGGTGTCCCTGCTCGAGGGTGAAGCTGTGTGAGTCAAT 107889
Qy 3121 cacaagctcctggacggctggtgggtcatcaggtaggagggccctctccatccagagca 3180
Db 107890 CACAAGCTCCTGGACGGGTGTGGTGTATCAGGTAGGAGGGCCCTCTCCATCCAGAGCA 107949
Qy 3181 cccatctgagtcagccccagccagagcaggggtgttttaagggatctggggtgactgtccct 3240
Db 107950 CCCATCTCAGTCAGCCCCAGCCAGCAGCGGGTGTATTAGGGATCTGGGTGACTTGTCCCT 108009
Qy 3241 gggactctgggttaagccactgccctctctggtgttagtttccatctcagtagcagggag 3300
Db 108010 GGGACTCTGGGTAAAGCACTGCCCTCTCTGGGCTTAGTTTCCATCTCAGTAGCAGGGAG 108069
Qy 3301 ggatgagccaccttgcctgtcttgggtggaatcgaatgtccttgccttgccttgccttgcct 3360
Db 108070 GAATGAGCCACCTTGCCTGTCTTGTGGGATCCCAATGTCTCTGTCCCAAGTGGGTGCAT 108129
Qy 3361 t 3361
Db 108130 t 108130

RESULT 5
AC004166 AC004166 275197 bp DNA PRI 20-JUL-2000
LOCUS Homo sapiens Chromosome 7 BAC Clone 239c10, complete sequence.
DEFINITION AC004166
ACCESSION AC004166
VERSION AC004166.12 GI:8887011
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 275197)
AUTHORS Ren,Q., Burian,D., Huang,E., Meadows,S., Korenberg,J. and Roe,B.A.
TITLE Homo sapiens Chromosome 7 BAC Clone 239c10
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 275197)
AUTHORS Burian,D.M. and Roe,B.A.
TITLE Direct Submission
JOURNAL Submitted (21-FEB-1998) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
REFERENCE 3 (bases 1 to 275197)
AUTHORS Burian,D.M., Ren,Q., Meadows,S., Huang,E., Korenberg,J. and
Roe,B.A.
TITLE Direct Submission
JOURNAL Submitted (01-JUL-2000) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
REFERENCE 4 (bases 1 to 275197)
AUTHORS Ren,Q., Burian,D., Huang,E., Meadows,S., Korenberg,J. and Roe,B.A.
TITLE Direct Submission
JOURNAL Submitted (20-JUL-2000) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
COMMENT On Jul 1, 2000 this sequence version replaced gi:8779478.
FEATURES
Location/Qualifiers
source 1.275197
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="7"
/clone="239c10"
BASE COUNT 70874 a 66648 c 66737 g 70938 t
ORIGIN

Query Match 98.0%; Score 3293.2; DB 9; Length 275197;
Best Local Similarity 99.3%; Pred. No. 0;

Matches 3339; Conservative 0; Mismatches 18; Indels 5; Gaps 3;				
Qy	1	aaaggtcaggcactgtggtcctacgtcctgtaatccacgcaactttggaggccgagcggt	60	
Db	43597	AAAAGGTCAGGCACATGTGGCTCATGCTGTAATCCAGCACATTTGGAGGCGCGAGCGGT	43656	
Qy	61	tgaatcaactgaggtcagaggttcagaccagcctgaccaatatcgtgaaactccatctc	120	
Db	43657	TGAATCACCTGAGGTCAGGAGTTCAGAGCCAGCCAGCTGACCAATATCGTGAATCCATCTC	43716	
Qy	121	tactaaaaatacaaaaatagccaggtgtgtgcggtgtgttagtccacgtactctg	180	
Db	43717	TACTAAAAATACAAAATTAGCCAGGTGTGTGGCGGTGTGTAGTCCAGCTACTTTG	43776	
Qy	181	ggaggtcaggcaggtgaattacttgaacctgggaggtgaggttgcaatgagccaagat	240	
Db	43777	GGAGGCTGAGGCAGGTGAATTACTTTGAACCTGGGAGGTGGAGTTGCAATGAGCCAAAGAT	43836	
Qy	241	tgcacctgcactccagtgcagagcgagactccatctcaaaaaaataaaaaaag	300	
Db	43837	TGCACCACTGCACCTCAGTGACAGAGCAGACTCCATCTC---AAAAAATAAAAAAAG	43893	
Qy	301	ttggaaaaagccaggtcagtggtccacgctgtaatcccaacactttaagaggtcga	360	
Db	43894	TTGGAAAAAGGCCAGGTGCAGTGGCTCCACGCTGTAATCCCAACACTTTAAAGAGGCTGA	43953	
Qy	361	ggtggagaatcctttgagcccgaggttcgagaccagcctgggcatgtgccagacct	420	
Db	43954	GGTGGGAGAACTCTTTGAGCCAGGAGTTTCGAGACCAGCTTGGGATTTGTCCTCAAGACCT	44013	
Qy	421	tgctttcaaaaaataatagccgggtgtgtgtggtacatctctgtgtccagctattcggg	480	
Db	44014	TGCTCTTACAAAAAATTAGCCGGGTGTGGTGATACGCTATGTGTCCAGCTATTTCGGG	44073	
Qy	481	aggtcgagccaggagattgctgaacctagaggtcttaggctgtagtgagctgtgac	540	
Db	44074	AGGCTGAGCGCAGGAGATTGCTTGAGCTAGAGTCCAGGGCTGTAGTGAGCTGTGTATCA	44133	
Qy	541	cgtcacttactctagctgggcaacagagcaagactctgtcccaaaaaaataaaa	600	
Db	44134	CGTCACTGTACTTAGCTTGGCCACAGACAGCAAGACTCTGTCTCCAAAAAAGAAATAAA	44193	
Qy	601	gttgggaaggctcaactcaactcatcagatgaaacaaagacatgtttgaagtgtgagcc	660	
Db	44194	GTTGGGAAGGCTCACTAACTTCATCAGATGAGAACAAAGGACATGTTTGAAGTGTGAGGC	44253	
Qy	661	cgaagcttggaagaactatcgcccaagaaaatgcagggcgagagactcaaatgcca	720	
Db	44254	CGAAGCCTGGAGAACGCTATGCGCCCAAGAAATGCAGGGCAGCAGAGACTCAAGATGCCA	44313	
Qy	721	gcgcctgtcttgaggccagatgggcccctgcaatgtcccactcacctgcctccctctt	780	
Db	44314	GCGCCTGTCTTGAGGCCACAGATGGGCCCTGCAATGCCCACTACCTGACCTGCTCCTCTT	44373	
Qy	781	gcccagacatacaocggcccccatcctcgcagactgaccgcgcattgcccactacag	840	
Db	44374	GCCCCAGACATACCGGGCCCCATCTCTGCAGAGCTACCGCGCCATTGCCAACTACGAG	44433	
Qy	841	aagacctcgagctccagatggtctctccacggggagcgtgtgtgaggtcgtgagaag	900	
Db	44434	AAGACCTCGGGCTCCGAGATGGCTGTGCCACGGGGAGCTGGTGGAGGTGTGTGAGAG	44493	
Qy	901	agcagagcggtcagacctcccaacttacggggtcccttcccctggtgtgtcagaaccca	960	
Db	44494	AGCGAGCGGTACAGACCTCCACCTTACGGGGCTCTTCCCTGGTGTCTCAGGNACCCA	44553	
Qy	961	cagccaagccccctgccagggtcagggcagcctggccccctggaggactccagctctg	1020	
Db	44554	CAGCCACAAGCCCCCTGCCAAGGCTCAGGCAGCCTTGCCCCCTGGGAGGACTCCGGCTCTG	44613	
Qy	1021	ttaggggcccataaatctctcccaactgtgggtcgctctctctcttagtgcacccct	1080	
Db	44614	TTAGGGGCCCTAAATGTCTCTCCACACACTGTGGGTGCGCTTCTGTCTTAGTGTGCACCT	44673	

Qy	1081	gtgggtggtgtgggcattctgtgcattgagccggggcggggcatgtctgctgtctgt	1140	
Db	44674	GTGGTGGCTGTGGGCATCTGTGCATGGCAGGCCGGGGCGGCATGTCTGCGCTTCTGT	44733	
Qy	1141	ctggatgggatagggaacgctctgttcattatgaaagtgggctcagagctgtgtattctgga	1200	
Db	44734	CTGGATGGGTATGGGACCGCTCTGTTTCATTATGAAGTGGGCTCAGAGCTGTGATCTGTGA	44793	
Qy	1201	gcattgtgcattgcattgcattgcattgcattgcattgcattgcattgcattgcattgc	1260	
Db	44794	GCATGTGTGCATGCATGCATGTGACCTCATTTCCAGTGTGTGAAGTGTGACATTTCCAA	44853	
Qy	1261	atctgagcattggacattcagttgtctgtccctgtgtccctcacatccctcgaaggctg	1320	
Db	44854	ATCTCAGCATTTGGACATCAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	44913	
Qy	1321	caggagcgcctgggcccctgccctcagtcacattcccgcacctctggcacaggttggcg	1380	
Db	44914	CAGGAGCGCTGGGCCCTGCCCTCAGTCACTATCCCGCACCTCTGGCACAGGTGGTG	44973	
Qy	1381	gttctgtcagatgaagcaaacgaggtgtggtccagcatctctcctcagaccctcga	1440	
Db	44974	GTCTGTGATGAAGCAAGCAGGCTGGATCCACGCTCTTCTCTGAGCCCTTGA	45033	
Qy	1441	cagtcctgacgagacgagaccctgagcccaactatgcaggtgccccctgcctcccgag	1500	
Db	45034	CAGTCTCTGACGAGACGGGAAGACCTTGAGCCCAACTATTCGAGGTGCCCTGCCCTCCGAG	45093	
Qy	1501	gctgtagggggtgtggggaagggggcagggggtcagggatattgtagtgcactcttg	1560	
Db	45094	GCTGTAGGGGTGTGGGAGAAAGGGGACGACAGGGCTCAGGGATATTGAGTGACTGCTTTG	45153	
Qy	1561	gagtcgtgggt	1620	
Db	45154	GAGTCTGGGCTGGT	45213	
Qy	1621	cttgggggccccctggcaggt	1680	
Db	45214	CTTGGGGCCCCCTGGCAGGTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	45273	
Qy	1681	agacactcgggaggtgtggggccagttgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt	1740	
Db	45274	AGACACTGGGAGAGTGGGAGGCCAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	45333	
Qy	1741	gaagctgaaggtgagcagaccttaggtcaggaagggggtcgtcgtggaagtgggggca	1800	
Db	45334	GAAGCTGAAGGATGAGCAGACCTTAGGCTCAGGAAGGAGGGCTGCTGGAAGTGGGGCA	45393	
Qy	1801	tcatcactgaccagaaagggaaaactggcagtgccagggcttgatggggcctgcattgag	1860	
Db	45394	TCATCACTGACCAGAAAGGGGAAACTGGCAGTGGCCAGGGCTGGATGGGGCTTCGATTGAG	45453	
Qy	1861	cttgaaaaaactataatagattgggttaccattttattttattttattttattttatttt	1920	
Db	45454	CTTGAAAAAACTATAATAGAAATGCTTACCATTATTTATTTATTTATTTATTTATTTATTT	45513	
Qy	1921	tactttttgagatagatgactcactcccttgcctaaggtggagtgcggtgggtgcctctc	1980	
Db	45514	TACTTTTTTGAATAGATGATCTCACCCTTGTCTAAGTGTGGAGTGGGTGGTGTATCTC	45573	
Qy	1981	agtcactgcacacctctcctcccaggaatcaagtgtatctccagcctcagcctccccagg	2040	
Db	45574	AGCTCACTGCACACCTCTGCCCTCCAGGATCAAGTGATTTCTCCAGGCTCAGGCTCCCCAAG	45633	
Qy	2041	tagctgggattacaagcatgcacacccatgctggataatttttttttttttttttttttt	2100	
Db	45634	TAGCTGGGATTACAAGCATGCACCACTGCCTGGCATAAATTTTGTATTTTGTAGTTGAGA	45693	
Qy	2101	cgggggttcca--ccaggttggccagactggtctcgaaactctcgaaactcaggggtgactgct	2159	
Db	45694	CGGGGTTTTACCCAGGTTGGCCAGACTGTCTCGAACTTCTGAACTTCTGACCTCAGGTGATCTGCT	45753	

QY 1 aaaggctcaggcactgtgtgctcatctcctgtataatccacagcaactttggaggccgagcggt 60
|||||
Db 10114 AAAAGTCAAGGACACTGTGGCTCATGCTTAATCCAGACACTTTGGAGGCGGAGCGGT 10173
|||||
QY 61 tgaatcacctgaggtcaggagttcagagaccacttgaccaataatcgtgaaactccatctc 120
|||||
Db 10174 TGAATCACTGAGGTTCAGAGTTCGAGACCAAGCCCTGACCAATATCGTAAATCCATCTC 10233
|||||
QY 121 tactaaaaacaaaaattagccaggtgtggtggcgqgtgctgttagtcccagctacttg 180
|||||
Db 10234 TACTAAAAATACAAAAATTAGCCAGGTGGTGGCGGTCTGTAGTCCCAAGCTACTTG 10293
|||||
QY 181 gaggctcaggcaggtgaattacttgtaacctgggaggtggaggttgcaatgagccaaagt 240
|||||
Db 10294 GGAGGCTCAGGCAGGTGAATTAATTAACCTGGGAGGTGGAGGTTGCAATGAGCCAAAGAT 10353
|||||
QY 241 tgcacctgcactccagtgcagagccagactcccatctcaaaaaaataaaaaaag 300
|||||
Db 10354 TGCACCACTGCACTCCAGTGACAGAGGAGACTCCATCTC - AAAAAAATAAAAAAAG 10412
|||||
QY 301 ttgggaaaggccaggtgcagttggctccacgcctgttaatcccaacactttaagaggtga 360
|||||
Db 10413 TTGGGAAAAGGCCAGGTGCAGTGGCTCCACGCTGTATATCCCAACACTTTAAGAGGCTCA 10472
|||||
QY 361 ggtggagaatcctttgagccaggagttcgagaccagcctgggcatgtgtcccaagaact 420
|||||
Db 10473 GGTGGGAGAACTCTTTAGCCCCAGGAGTTGAGAGCAGCCTGGGCATTTGCCAAGACCT 10532
|||||
QY 421 tctcttcaaaaaattagccaggtgtggtggcatacgtctgtgtccagctattcggg 480
|||||
Db 10533 TGTCTTTACCAAAATAGCCGGGTGTGGTGCACTACGTCTATGGTCCAGCTATTCCGGG 10592
|||||
QY 481 aggctgagcaggagattgcttggacctagagttcctagggctgtagtgagctgtgata 540
|||||
Db 10593 AGGCTGAGCAGGGAGATTGCTTTGAGCTAGGAGTCCAGGGCTGTAGTGAGCTGTGATCA 10652
|||||
QY 541 cgtcactgtactctagcctgggcaacagagcaagactctgttcccaaaaaaataaaa 600
|||||
Db 10653 CGTCACTGTACTCTAGCTGGGCAACAGAGCAAGACTCTGTCTCCCAAAAAGAAATAAA 10712
|||||
QY 601 gttggaaaggtcactaacttcacagatgagaaacaagacatgtttgaaagtggagcc 660
|||||
Db 10713 GTTGGAAAGGCTACTAATCTCATGATGAGAACAGGACATGTTTGAAGTGTGAGGC 10772
|||||
QY 661 cgaagcctggagaaacgctatgcgccagagaaatgcagggcagcagagactcaagatgcca 720
|||||
Db 10773 CGAAGCCTGGAGAACGCTATGCGCCACAGAAATGCAGGGCAGCAGAGACTCAAGATGCCA 10832
|||||
QY 721 gcgcctgttctggagggccagatgggcccctgcaatgcccactcaacctgcccctctt 780
|||||
Db 10833 GCGCCTGTCTGGAGGCCAGATGGGCCCTGCAATGCCACCTCACCCCTGCCCTCCTCTT 10892
|||||
QY 781 gcccagacatcacggcccatactcctgcagacgtaccgcgccattgcgcactacag 840
|||||
Db 10893 GCCCAGACATCACCGCCCCATATCTCTGCAGAGCTACCGGCCATTGCCAACTACGAG 10952
|||||
QY 841 aagacctgggctccagatggctctgtccacgggggacgtgtggtgaggtcgtggagag 900
|||||
Db 10953 AAGACCTCGGGCTCCAGATGCTCTGTCCACGGGGAGCTGTTGGAGGTCTAGAGAG 11012
|||||
QY 901 agcagagcgggtcagaacctcccaaccttacggggctccttcccctgggtcgtcaggaacca 960
|||||
Db 11013 AGCAGAGCGGTTCAGACCTCCCACTTACGGGGCTCTCTTCCCTGGTCTCAGGAACCCA 11072
|||||
QY 961 cagccacaagccccctgccaaaggtcaggcaacctggcccctggaggagactccacctgtg 1020
|||||
Db 11073 CAGCCACAAGCCCCCTGCAAGCTCAGGACGCTTGGCCCCCTGGGAGGACTCCGGCTCTG 11132
|||||
QY 1021 ttaggggccctaaatgtctctcccaacctgtgggtcgcccttctctcttagtgcacct 1080
|||||
Db 11133 TTAGGGGCCCTAAATGTCTCTCCCACTAGTGGGTGGCTTCTCTCTTAGTGTGACCCCT 11192
|||||
QY 1081 gtggtggtgtgggcatctgtgcatggcaggccggggcggggcatgtctgtctgttctgt 1140
|||||

Db 11193 GTGGTGGCTGTGGGATCTGTGCATGGCAGCGCGGGCGGCATGCTGCGTGTCTGT 11252
|||||
QY 1141 ctggatgggtatggacacgtctgttcattatgaagtgggctcagagctgtgattctgtga 1200
|||||
Db 11253 CTGGATGGGTATGGGACCGCTGTTCATTATGAAGTGGGCTCAGAGCTGTGATCTGTGA 11312
|||||
QY 1201 gcatgtgtcatgtcatgtcagctcattgtccagttgtgtgaagggtgacatttccaa 1260
|||||
Db 11313 GCATGTGTGCATGTCATGTGACCTCATTTGCCAGTGTGTGAAGTGCACATTTCCAA 11372
|||||
QY 1261 abctgagcaattggacatacagttgtgtgtccctgtgttctccacatccctcgtatggctg 1320
|||||
Db 11373 ATCTGAGCAATTGGACATCAGTGTGTGTCTGCTCTGCTCCTCACCATCCCTGATGGCTG 11432
|||||
QY 1321 caggagccgctggqccctgcccctcagtcacatctccgcacactctggtcacaggttgggtg 1380
|||||
Db 11433 CAGGAGCCGCTGGGCCCTGCCCCCTCAGTCACATTTCCCGACCTCTGGCACAGTTGGTG 11492
|||||
QY 1381 gttctgtcagatgaaagcaagcgaggtgtggtatccacagcatccttctcgagccccctgga 1440
|||||
Db 11493 GTTCTGTCAAGTGAAGCAAAAGCAGGCTGGATCCCAGCATCTTCTCGAGCCCTGGA 11552
|||||
QY 1441 cagtcctgcagacggaagacccctgagccccactatgcaggtgccccctgccccctcag 1500
|||||
Db 11553 CAGTCTCGACGAGACGGAAGACCTTGAGCCCAACTATGAGAGTGCCTCCTGCCCTCCGAG 11612
|||||
QY 1501 gctgtaggggtgtggagaaaggagcagcagggctcagggatattagtgactgcttctg 1560
|||||
Db 11613 GCTGTAGGGGTGTGGGAAAGGGCAGGCGTCTGAGGATATTGAGTGACTGCTTTTG 11672
|||||
QY 1561 gsgtctgggctgtgtgtgctgtggcagaaagtcagggtcagggtcctaagatctcatcggtcctg 1620
|||||
Db 11673 GAGTCTGGGCTGTGTGCTGACTTTGGCAGAAAAGTCAGGCTAAGATCTCATCGACTCTGG 11732
|||||
QY 1621 cttgggggccccctggcaggtgtgtgtgtcccttggctggacagggaaacccggggaggagc 1680
|||||
Db 11733 CTTGGGGGCCCTGGCAGGTTGTGTGCCCTTGGTCTGGACAGGGAA - CAGGAGGAGGAGC 11791
|||||
QY 1681 agacactcgggagtgaggggcagtggtgtctgtgtgatatgtggcaggtctcagttg 1740
|||||
Db 11792 AGACCACTGGGAGAGTGGGAGGCCAGTGTGTGTGTGGATATGTGGCCAGGTTCAAGTGG 11851
|||||
QY 1741 gaagctgaaggatgagcagaccttaggctcaggaaaggaggtgctgctgaaagtgggggca 1800
|||||
Db 11852 GAAGCTGAAGGATGAGCAGACCTTAGGCTCAGGAAGAGGGCTGCTGGAAGTGGGGCA 11911
|||||
QY 1801 tcatcactgacagaaaaagggaaaactggcagtgccagggtggatggggccctgcattgag 1860
|||||
Db 11912 TCATCACTGACCAGAAAAGGAAACTGGCAGTGCACAGGCTGGATGGGCGCTGCATTGAG 11971
|||||
QY 1861 cttgaaaaaactataatagaattggttaccattttattttattttattttattttattt 1920
|||||
Db 11972 CTTGAAAAAAACTAATAAGAAATGGTTACCATTTTCATTTTATTTATTTATTTATTT 12031
|||||
QY 1921 tactttttgagatagagctcactccttgcctaaagcctggaggtgcgtggtgctatctc 1980
|||||
Db 12032 TACTTTTTTGTAGATAGAGTCTCAC'TCCCTTGCTAAGG - TGGAGTGGGGTGGTGTATCTC 12090
|||||
QY 1981 agctcactgcaacctctgcctcccaggatcaagtgatcttcacagctcagcctccccagg 2040
|||||
Db 12091 AGCTCACTCAACACTCTGCCTCCAGGATCAAGTGATTTCTCCAGGCTCAGCCTCCCCAGG 12150
|||||
QY 2041 tagctgggattacaagcatgcccaccatgcctgctgataatttttatttttagttgaga 2100
|||||
Db 12151 TAGCTGGGATTAACAGCATGCACCACTACCATGCTCTGGATAATTTTGTATTATTTAGTTGAGA 12210
|||||
QY 2101 cgggggttccacaggttggccagactgggtcctcgaaactcttcgactcaggtgactgcctg 2160
|||||
Db 12211 CGGGGTTTCCACAGGTTGGCCAGACTGGTCTCGAACTTCTGACCTCAGGTGATCTGCCTG 12270
|||||
QY 2161 cctcgccctcccaagtgctggaattacaagatgtgagccactgtccctggcctggttacc 2220
|||||

Db 12271 CCTCGCCTCCCAAAAGTCTGGAATTACAGATGTGAGCCACTGTCCCTGGCGTGTACC 12330
Qy 2221 cacattttaaatgagtgatttcaaccttttatatgtgaatttacaagctgttttttttt 2280
Db 12331 CACATTTTAAATGAGTGATTTTCAACCTTTTATGTGATTTACAGCTTG--TTTTTTT 12388
Qy 2281 ttttttgacaaaagtctggctctgtcccccaggctggtgagtgagtaaatgcaatctcagc 2340
Db 12389 TTTTTTGTAGACAAAGTCTGTTCTCTCACCCAGGCTGGAGTGCAGTAATGCAATCTCAGC 12448
Qy 2341 tcaactgaacctagctctcctctgggttcaagcaattctcctgcctcagccaactgagtagc 2400
Db 12449 TCACCTGCAACCTTAGCTCTCTGGTTCGAAGCAATTTCTCTGCTCAGCCACCTGAGTAG- 12507
Qy 2401 ctggggttacaggcatgacacacagccagcgaactaatttttttttttttttttttttt 2460
Db 12508 CTGGGATTACAGGCATGACCAACCCAGCCAGCTAAATTTTGTATTTTGTAGAGATG 12567
Qy 2461 gggtttcgcattgttggccaggctggtctcgaacctcctgacctcagtgatccgcgcgcc 2520
Db 12568 GGGTTTCCCATGTTGGCCAGGCTGCTCGAACTCTCTGACCTCAGTGCAGTCCGCCGCC 12627
Qy 2521 ttggcctcccaaatgctagattacaggttggaacaccttgccaccctgtggtctatc 2580
Db 12628 TTGGCCTCCAAAGTCTAGGATTACAGTGGGAACCACTCGCCAGCCTGTGGCTATC 12687
Qy 2581 gtttaaacactgggaagcctgcagccccccagccagtcagttagctgcagctgagcagtt 2640
Db 12688 GTTTAAACACTGGGAAGGCTGACGCCCCAGCCGACAGTTAGTGCAGCTGACGAGTT 12747
Qy 2641 ccagtgccagtagcagtagctcaccacctactcatggtgctgattcttctgtcatagt 2700
Db 12748 CCCAGTGCAGTAGCAGGATGCTCCACCCTACTCATGGCTGATCTCTGTCTAGT 12807
Qy 2701 gaagtctctggacagaccttcattgttggtatctctggtcccccagagtggtggaat 2760
Db 12808 GAAGTCTCTGGACAGACCTTCACTGTTATGGGATCTCTGGTCCCGAGAGTGGTGGCAAT 12867
Qy 2761 gaatggagtgacaagctcacctgggtgtaggggcagagggccgaagtcacagagtata 2820
Db 12868 GAATGGGAGTGGACAAAGCTCACCTGGGTGTAGGGGGCAGAGGGCCGAAGTCCAGAGTGTA 12927
Qy 2821 cccccagagtggtgtccagcagagagcttgccagggatctctggtggagcagaggggtgg 2880
Db 12928 CCCCCAGAGTGGTCCAGCAGAGCTTCCCGAGGGATCTGGGATGGAGCAGAGGGGTGG 12987
Qy 2881 agggggagaccacagaagagggggaactgtggtccctgggtgggtctctggagtgccctggag 2940
Db 12988 AGGGAGGAGACCCAGAAGAGGGGGAACCTGTGGGCCCTGGGTGGGTCTGGAGTGGCGAG 13047
Qy 2941 gaagcccagggcagagagagagagatgggatgggtggtgcgagccccagagctgggcccagc 3000
Db 13048 GAAGCCAGGCCAGAGAGAGAGATGGGATGGGTGGCGAGCCCGCCAGGCTGGGCCGACC 13107
Qy 3001 tcacactgtctctgtgcccctgcggtgacacaggtgagccatacgtgcctatcaaggcc 3060
Db 13108 TCACACTGTGCTCTGTGCCCTGCGCTGGACAGGTGAGCCATACGTGCGCCATCAAGGCC 13167
Qy 3061 tcaactgtctgtgaggggagacaggtgtccctgtctcgaggggtgaagctgttgaggttaatt 3120
Db 13168 TACACTGTGTGGAGGGGACAGAGTGTCCCTGCTCGAGGGGTGAAGCTGTTGAGGTCAAT 13227
Qy 3121 cacaagctctgagcgtgtggtgtcatcaggtaggagggccctctccatccagagaca 3180
Db 13228 CACAAGCTCTCGACGGCTGTGGTGTATCATAGTAGGAGGGCCCTCTCCATCCAGAGCA 13287
Qy 3181 cccatctgagtcagccccagcagcaggggtgttttagggatctgggtgactgtgcct 3240
Db 13288 CCCATCTAGTCAGCCCCAGCCAGGAGGCGGTGTTAGGGATCTGGGTGACTTGTCCCT 13347
Qy 3241 gggactctgggttaagccactgcctctctctgtgggttagtttccatctcagtagcagggag 3300
Db 13348 GGGACTCTGGGTAAAGCCACTGCCCTCTCTGGGCTTAGTTTCCATCTCAGTAGCAGGGAG 13407

Qy 3301 ggatgagccacccttgctctgttggtggtatccaatgtccttgccaagtgggtgcat 3360
Db 13408 GAATGAGCCACCCCTTGCTGTCTGTGGGATCCAATGCTCTTGTCCAAGTGGTGCAAT 13467
Qy 3361 t 3361
Db 13468 T 13468
RESULT 7
HS47P43S02 3348 bp DNA PRI 07-JAN-1998
LOCUS Homo sapiens p47-phox (NCF1) pseudogene, clone P43, exons 6-8.
DEFINITION U60971
ACCESSION U60971 GI:2754715
VERSION U60971.1
KEYWORDS A p47-phox pseudogene carries the most common mutation causing
SEGMENT p47-phox- deficient chronic granulomatous disease
SOURCE J. Clin. Invest. 100 (8), 1907-1918 (1997)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 3348)
AUTHORS Gorlach, A., Lee, P.L., Roesler, J., Hopkins, P.J., Christensen, B.,
Green, E.D., Chanock, S.J. and Curnutte, J.T.
TITLE A p47-phox- deficient chronic granulomatous disease
JOURNAL J. Clin. Invest. 100 (8), 1907-1918 (1997)
MEDLINE 97474758
REFERENCE 2 (bases 1 to 3348)
AUTHORS Chanock, S.J., Roesler, J., Hopkins, P.J., Lee, P.L., Bassett, D.T.,
Christensen, B., Curnutte, J.T. and Gorlach, A.
TITLE Genomic structure and the identification of multiple polymorphisms
in the p47-phox gene
JOURNAL unpublished
REFERENCE 3 (bases 1 to 3348)
AUTHORS Gorlach, A., Lee, P.L., Roesler, J., Chanock, S.J. and Curnutte, J.T.
TITLE Direct Submission
JOURNAL Submitted (17-JUN-1996) Molecular and Experimental Medicine, The
Scripps Research Institute, 10550 N. Torrey Pines Rd., CAL 1, La
Jolla, CA 92037, USA
FEATURES
source
location/Qualifiers
1..3348
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="P1 clone P43"
/chromosome="7"
/map="7q11.23"
/note="Human genomic P1 clone P43 is a p47-phox pseudogene
which is defined by the presence of a GT deletion at the
beginning of exon 2"
507..629
/gene="NCF1"
exon
pseudo
/number=6
1093..1200
/gene="NCF1"
pseudo
/number=7
2749..2866
/gene="NCF1"
pseudo
/number=8
BASE COUNT 715 a 854 c 1003 g 776 t
ORIGIN
Query Match 89.1%; Score 2993.2; DB 9; Length 3348;
Best Local Similarity 99.1%; Pred. No. 0;
Matches 3052; Conservative 0; Mismatches 23; Indels 5; Gaps 4;
Qy 282 aaaaaaaaaaaaaattgggaaagggccaggtgcagtggtccacgcctgtaatcc 341
|||||

Db 1 AAAAAAAAAAAAAAAAAAGTTGGAAAAAGCCAGGTACAGTGGCTCCACGCCGTGTAATCC 60
Qy 342 caacacttaagagctgagtgaggagaatctcttgagccagagagttcgagacagcct 401
Db 61 CAACACTTTAGAGCGCTGAGGTGGGAGAACTCTTTGAGCCCGAGGATTCGAGACAGCCT 120
Qy 402 gggcattgtcccaagacctgtctttacaaaaaattagcccgggtgtgtgtgacatcagctct 461
Db 121 GGGCATTTCCCAAGACCTTGTCTTTACACAAAATTAGCCGGCTGTGTGGCATACGTCCT 180
Qy 462 gtgtcccaagctattccggagagctgaggcaggagagatgtgtgagcctaggagctcaggg 521
Db 181 GTGTGCCAGCTATTTCGGGAGGGCTGAGCGAGGGAGATTGCTTGAGCCCTAGGAGTCTAGG 240
Qy 522 ctgtagtgagctgtgatcaactgactgtactctagctcgtgggcaacagagcaagactctgt 581
Db 241 CTGTAGTGAGCTGTGATCACGCTACTGTACTCTAGCCTGGGCAACAGAGCAAGACTCTGT 300
Qy 592 ctccaaaaaagaaataaagtgtggaaaaggtctcaactaacttcacagatggaacaaaga 641
Db 301 CTCCCAAAAAGAAATAAAGTTGGGAAAGGCTCACTAACTTTCATCAGATGAGAACAAAGA 360
Qy 642 catgtttgaagtgtgagccgaagcttgagaaagctgagaaagctatgcgccagaaatgcaggca 701
Db 361 CATGTTTGAAGTGTGAGGCCCGAAGCCTGGGAACGCTATGCGCCCGAGAAATGCGAGGCA 420
Qy 702 gcagagactcaagatgcagagcgctgttctgaggccagagatgggcccctgcaatgcccac 761
Db 421 GCAGAGACTCAAGATGCCAGCGCCTGTCTCGAGGCCCAGATGGGCCCTGCAATGCCAC 480
Qy 762 tcaacctgcctccctctgtgccccagacataccaggcccccacatccctgcagagatccg 821
Db 481 TCACCCCTGCCCTCCTCTGTGCCCGACAGATCACCGCGGCCCATCATCCTGCGAGCTACCG 540
Qy 822 cgcattccagactacagagaagacctcgggtccagatggtctgtccagcggggagagt 881
Db 541 CGCCATTGCCAACTACGAGAAGACTCGGGCTCCGAGATGGCTCTGTCCACGGGGAGCT 600
Qy 882 ggtgaggtcgtgagaaagcagagacggtcagacacctccacacctagcgggctccttcc 941
Db 601 GGTGAGGTCTGGNAGAGAGCGGAGCGGTGAGACCTCCCACTTACGGGGCTCCTTCC 660
Qy 942 cctggtgtcaggaaaccacagccacaaagccccctgccaaggtctcagcagcctggcccc 1001
Db 661 CCTGTGTCTAGGAACCCACAGCCACAAAGCCCTCTGCCAA-GCTCAGGACGCTGGCCCC 719
Qy 1002 tgggaggaactccagctcgtttaggggcccataaagtctctcccacactgtgggtgcctt 1061
Db 720 TGGGAGGACTCAGGCTCTGTGTAGGGGCCCTAAATGTCTCTCCCACTGTGGGTGCGCTT 779
Qy 1062 ctctcttagtgtcacacctgt 1121
Db 780 CTCCTTAAAGTGTGNACCCCTGTGTGACTGTGGGCATCTGTGCATGGCAGCGGGGGCGG 839
Qy 1122 gcatgtctgcgtgttctgtctggtatgggtatgggaccgtctgttcaattatgaagtggct 1181
Db 840 GCATGCTGCGGTGTCTGTCTGTGGATGGGTATGGGACCGTCTGTTCATTATGAAGTGGCT 899
Qy 1182 cagagctgtattctgtgagacatgtgtcatgtcatgtcatgtcaacctatgttcaagtgtg 1241
Db 900 CAGAGCTGTGATTTCTGTGAGCATGTGTGCATGTCATGTGACCTCATTTGTCCAGTGTG 959
Qy 1242 gtgaaggtgacatttccaaatctgaagattggacatgaattgtgtgtgtgtgtgtgtgtgtgt 1301
Db 960 GTGAAGGTGACATTTCCAAAATCTGAGCAATTGGACATCATGATGTGTGTGACCCCTGTGCC 1019
Qy 1302 tcaccatccctgtatgtgtcagggagccgctgggcccctgcccctcaagtccattctccgca 1361
Db 1020 TCACCATCCCTGATGGCTGCGAGGAGCGCTGGGCCCTGCCCCCTCAGTCACATTCGCCGA 1079
Qy 1362 cctctggcacaggttgggtgtctgtctcagatgaaagcaaaagcagggttgatccacagcat 1421
Db 1080 CCTCTGGCACAGGTGTGTGGTGTCTGTGATGATAAGCAAGCGAGGCTGGATCCCGACAT 1139

Qy 1422 ccttctcagccccctggacagtcctcagagaaagacccctgagccccactatgcag 1481
Db 1140 CCTTCTCGAGCCCTGCACAGTCTCTGACGAGAGAGACCCCTGAGCCCAACTATGCAG 1199
Qy 1482 gtgccccctgcctcccgaggtgtaggggtgtgaggagaaagggcaggaggtcaggg 1541
Db 1200 GTGCCCTTGCCTCCGAGGCTGTAGGGGTGTGGGAGAAAGGGGCGAGCGCTCAGGG 1259
Qy 1542 atattgagtactcttggagctgtgggctgttctcagcttgagaaagctcaggct 1601
Db 1260 ATATTGAGTGACTGTCTTGGAGTCTGGGCTGTGCTGGCTTGGCAGAAAGCTCAGGGCT 1319
Qy 1602 aagactctcatcggctcgtccttgggggcccctggcaggtgtgtatcccttggctcggaca 1661
Db 1320 AAGATCTCATGCTCTGCTTGGGGCCCTGGCAGGTTGTGATGCCCTTGGCTCGGACA 1379
Qy 1662 gggaaaccaggaggagcagacgactcggagagatgggagggccaggtgtgtctgtgat 1721
Db 1380 GGGAA-CAGGAGGAGGAGCAGACGACTCGGGAGAGTGGGAGGCCAGTGTGTCTGTGGAT 1438
Qy 1722 atgtggccaggttcagtggaagctgaagatgagcagacaccttaggctcagaaagaggg 1781
Db 1439 ATGTGCCCAGGTTTCACTGGGAGCTGAAGATGACAGACCTTAGGCTCAGGAAGAGGG 1498
Qy 1782 ctgcttgaagtgggggcatcatcactcagcagaagggaaaaactggcagtgccagggct 1841
Db 1499 CTGCCCTGGAAGTGGGGGCATCATCTGACCAGAAAGGAAACTGGCAGTGCCAGGGCT 1558
Qy 1842 ggaatggggccctgcattgagcttgaaaaaactataatagaattgtgtaccattttat 1901
Db 1559 GGATGGGGCTGCATTTGAGCTTGAAAAAACTATTAATAGAATTGGTTACCAATTTCAATTT 1618
Qy 1902 attatttatttatttatttatttatttatttatttatttatttatttatttatttattt 1961
Db 1619 ATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATT 1678
Qy 1962 agtgggtgtgtgtatctcagctcactgcaacctgtcctcccaggatcaagtgtattctc 2021
Db 1679 AGTGGGGTGGTCTATCTCAGCTCAGCTCACTGCAACCTCTGCCCTCCCAGGATCAAGTATT 1738
Qy 2022 cagctcagctcccccaggtagctgggattacaagcatgcaccaccatgcctcggataatt 2081
Db 1739 CAGCCTCAGCCTCCCGAGGTAGTTGGGATTACAAGCATGACAGCATGCGCTGGATAATT 1798
Qy 2082 tttgtatttttagttgagacgggtttcaccaggttgccagacgtgtctcgaactctcg 2141
Db 1799 TGTGTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATT 1858
Qy 2142 acctcaggtgatctcctgcctcggcctcccacaaagtgtcggaaatlacagatgtgagccac 2201
Db 1859 ACCTCAGGTGATCTGCCCTGGCTTGGGCTCCCAAAAGTGTGGAAATACAGATGTGAGCCAC 1918
Qy 2202 tgtccctggcctgtgtaccacattttaaaatggaggtgatttcaccccttttatgtggatt 2261
Db 1919 TGTCTCGCTGGCTGGTTACCCACATTTTAAATGGAGTGTATTTACCCCTTTTATGTGGATT 1978
Qy 2262 tacagcttgttt 2321
Db 1979 TACAGCT--TTTTTTTTTTTTTTTTTTTGGAGACAAAGTCTGGTCTGTACCCAGCGCTGGAGT 2036
Qy 2322 gcagtaatgcgaactcagctcactcgaaccttagcctcctcctcctcctcctcctcctcctc 2381
Db 2037 GCAGTAATGCAAATCTCAGTTCTACTCAACCTTAGGCTCTCTGGGTTCAAGCAATTTCTCTG 2096
Qy 2382 cctcagccaccttagtagctcgtgggttacaggagcatgcaccacagccagggcctaattttt 2441
Db 2097 CCTCAGCCACCTGAGTAG-CTGGGATTACAGGCATGACACCAACACCGCCAGGCTAATTTTT 2155
Qy 2442 tgtaatttttagtagagatgggggtttcgcgaattgtggcaggtgtgtctcgaactcctgac 2501
Db 2156 TGTATTATTTAGTAGAGATGGGGTTTCGCCATGTTGGCCAGGCTGTGTCTCGAACTCTCTGAC 2215

Qy 2502 ctccagtgatccgcccgccttggccctcccaagtgtcattagattacagatggaaacacct 2561
|||||
Db 2216 ctccagtgatccgcccgccttggccctcccaagtgtcattagattacagatggaaacacct 2275
Qy 2562 tgcacagcctgtggctatcgttttaaacactgggaaggcctcagccccagcgagcagt 2621
|||||
Db 2276 cgcgccagcctgtggctatcgttttaaacactgggaaggcctcagccccagcgagcagt 2335
Qy 2622 tagctcagctgagcagttccacgtccaggtagacgtagatctccacccactactatg 2681
Db 2336 tagctcagctgagcagttccacgtccaggtagacgtagatctccacccactactatg 2395
Qy 2682 gctgctctctgtcctatagtgagtgatgagtgagtgagtgagtgagtgagtgagtg 2741
|||||
Db 2396 gctgctctctgtcctatagtgagtgatgagtgagtgagtgagtgagtgagtgagtg 2455
Qy 2742 cccacagtggtggcaaatgaatggagtgagtgagtgagtgagtgagtgagtgagtg 2801
Db 2456 cccacagtggtggcaaatgaatggagtgagtgagtgagtgagtgagtgagtgagtg 2515
Qy 2802 ggcgaagtccagagtgatcccccagagtggtggcagcagagagcttccagagagatctg 2861
Db 2516 ggcgaagtccagagtgatcccccagagtggtggcagcagagagcttccagagagatctg 2575
Qy 2862 gcatgagcagagtggtgagtgagtgagtgagtgagtgagtgagtgagtgagtgag 2921
Db 2576 gcatgagcagagtggtgagtgagtgagtgagtgagtgagtgagtgagtgagtgag 2635
Qy 2922 ggtctgagtgcttggaggagccagcgccagagagagagagagagagagagagagag 2981
Db 2636 ggtctgagtgcttggaggagccagcgccagagagagagagagagagagagagagag 2695
Qy 2982 gcccagtggtggcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 3041
Db 2696 gcccagtggtggcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 2755
Qy 3042 atactgcgcatacagagcctacactgtgtgaggggagcagagtggtccctgtctgaggg 3101
Db 2756 atactgcgcatacagagcctacactgtgtgaggggagcagagtggtccctgtctgaggg 2815
Qy 3102 tgaagctgttgaagtgatcacaagctcctggagcagcagcagcagcagcagcagcag 3161
Db 2816 tgaagctgttgaagtgatcacaagctcctggagcagcagcagcagcagcagcagcag 2875
Qy 3162 cccctctccatccagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 3221
Db 2876 cccctctccatccagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 2935
Qy 3222 tctgggtgacttgcctgggactctgggtaagccactgcccctctctggtggttagttt 3281
Db 2936 tctgggtgacttgcctgggactctgggtaagccactgcccctctctggtggttagttt 2995
Qy 3282 ccactcagtaagcagggagggatagccaccccttgcctgtcttctgtggggtccaatgac 3341
Db 2996 ccactcagtaagcagggagggatagccaccccttgcctgtcttctgtggggtccaatgac 3055
Qy 3342 cttgtccaaagtgggtgcatt 3361
Db 3056 cttgtccaaagtgggtgcatt 3075

RESULT 8
LOCUS HS47P40S02 3350 bp DNA PRI 07-JAN-1998
DEFINITION Homo sapiens p47-phox (NCFL) pseudogene, clone P40, exons 6-8.
ACCESSION U61239
VERSION U61239.1 GI:2754722
KEYWORDS
SEGMENT 2 of 3
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 3350)
Gorlach,A., Lee,P.L., Roesler,J., Hopkins,P.J., Christensen,B.,
Green,E.D., Chanock,S.J. and Curnutte,J.T.
A p47-phox pseudogene carries the most common mutation causing
p47-phox- deficient chronic granulomatous disease
J. Clin. Invest. 100 (8), 1907-1918 (1997)
97474758
2 (bases 1 to 3350)
Chanock,S.J., Roesler,J., Hopkins,P.J., Lee,P.L., Bassett,D.T.,
Christensen,B., Curnutte,J.T. and Gorlach,A.
Characterization of the genomic structure of the p47-phox gene
Unpublished
3 (bases 1 to 3350)
Gorlach,A., Lee,P.L., Roesler,J., Chanock,S.J. and Curnutte,J.T.
Direct Submission
Submitted (19-JUN-1996) Molecular and Experimental Medicine, The
Scripps Research Institute, 10550 N. Torrey Pines Rd., CAL 1, La
Jolla, CA 92037, USA
Location/Qualifiers
1..3350
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="P40"
/chromosome="7"
/map="7q11.23"
/note="Human genomic P1 clone P40 is a p47-phox pseudogene
which is defined by the presence of a GT deletion at the
beginning of exon 2"
507..629
/gene="NCFL"
/pseudo
/number=6
1093..1200
/gene="NCFL"
/pseudo
/number=7
2749..2866
/gene="NCFL"
/pseudo
/number=8
BASE COUNT 711 a 856 c 1008 g 775 t
ORIGIN
Query Match 89.0%; Score 2990; DB 9; Length 3350;
Best Local Similarity 99.0%; Pred. No. 0;
Matches 3050; Conservative 0; Mismatches 25; Indels 5; Gaps 4;
Qy 282 aaaaaaaaaaaaaagtgtggaaaagggcaggtgcagtggtccacgcgtgtaatcc 341
Db 1 aaaaaaaaaaaaaaaaaagtgtggaaaagggcaggtgcagtggtccacgcgtgtaatcc 60
Qy 342 caacacttaagagctaggtggagagatcctttgagcccccagaggttcgagaccagct 401
Db 61 CAACACTTAAAGAGCTGAGGTGGAGAGATCCTTTGAGCCCCAGGAGTTCGAGACAGGCT 120
Qy 402 gggcattgtcccaagacctgtctttacaaaaaatagccgggtgtgtggtggtacacgtct 461
Db 121 GGGCATTGTCCCAAGACCTGTCTTTACCAGAAATTAGCCGGGTGTGGCATACGCT 180
Qy 462 gtgtccacgtattcgggaggtcagcagggagatgtcttgagcctaggagctctaggg 521
Db 181 GTGTGTCAGGCTATTTCGGGAGGCTGAGGAGGAGATGCTTTGAGCCTTAGGAGTCTAGGG 240
Qy 522 ctgtagtgagctgtagcagctcactgtactgtactgtactgtactgtactgtactgtact 581
Db 241 CTGTAGTGAGCTGTGATCAGCTACTGTACTGTACTGTACTGTACTGTACTGTACTGTACT 300
Qy 582 ctccaaaaaagaaaaaagtgtggaaaagggctcactaacctcatcagatgagaaacaaaga 641
Db 301 CTCCCAAAAAGAAAATAAAGTTGGGAAAGGCTCACTAACTTCATCAGATGAGAACAAAGA 360
Qy 642 catgtttgaagtgtgagccgaagcctggagaaacgtatgcgccccagggaaatgcaggggca 701

|||||
Db 361 CATGTTTGAAGTGTGAGGCCGAAGCCCTGGAGAACGCTATGCGCCCGAGAAATGCGAGGCA 420
Qy 702 gcagagactcaagatgcagcgccctgttctgtgagccagatgggcccctgcaatgccac 761
Db 421 GCAGAGACTCAAGATGCCAGCGCCTGTTCTGAGGCCCGAGATGGCCCTGCAATGCCAC 480
Qy 762 tcaacctgcccctcctcttgcccagacatcaacggcccccatcctcgagagctaccg 821
Db 481 TCACCTGCCCCCTCTCTGCGCCCGAGACATCACGGGGCCCCATCATCTCTGCGAGACTACCG 540
Qy 822 cgcattgcgcactacagaagacctcgggctccagatggctctgttccacggggagct 881
Db 541 CGCCATGCCAACTACGAGAAGACTCGGGCTCCGAGATGGCTCTGTCCACGGGGGAGCT 600
Qy 882 ggtggagctgtgagaagacgagagcggtcagacctcccaccttacggggtccctccc 941
Db 601 GGTGGAGTCTGTGAGAGACCGGAGCGGTTCAGACCTCCCACTTAGGGGACTCCTTCC 660
Qy 942 cctggtgtcaggaaacacagcccaagccccctgcaaggtcagcagcctggcccc 1001
Db 661 CTTGTTGCTCAGGAACCCACAGCCACAAAGCCCCCTGCCCCA--GCTCAGGCGAGCCTGGCCCC 719
Qy 1002 tggagagactcagctctgtttaggggcccctaaatgtcctcccacactgtgggtgcgctt- 1061
Db 720 TGGGAGGACTCCGGCTCTGTGTAGGGGGCCCTAAATGTCTCTCCCCACACTGTGGGTGCGCCTT 779
Qy 1062 ctctcttagtgcacccctgtggtggtgtgggcatctgtgcatggaagcgcggggcg 1121
Db 780 CTCCTTAGTGTGACCCCTGTGTGGTGTGGGCATCTGTGCATGGCAGCGCGGGCGGG 839
Qy 1122 gcatgtctgcgttctgtctgtggatgggta tgggacgctgtgttcattatgaagtgggct 1181
Db 840 GCATGCTCTGCGCTGTCTCTCTGGATGGGTATGGGACCGCTCTGTTCATTTATCAAGTGGGCT 899
Qy 1182 cagagctgtgattctgtgagcatgtgtgcatgcatgcatgcatgcatgcatgcatgcatg 1241
Db 900 CAGAGCTGTGTATCTGTGTGAGCATGTGTGCATGTGCATGTGCATGTGCATGTGCATGTG 959
Qy 1242 gtgaaggtgacatttcaaaatctgagcatgtgacatcagtgctgtgtctgtctgtctgtct 1301
Db 960 GTGAAGGTGACATTTCCAAATCTCAGCATTTGGACATCATAGTGTGTGTGTGTGTGTGTGT 1019
Qy 1302 tcacatccctctga tggctgcaggagcgctgggccccctcagtcacatcccgca 1361
Db 1020 TCACCATCCCTGATGGCTGCAGGAGCGCTGGCCCTGCCCCCTCAGTCACTTCCCGCA 1079
Qy 1362 cctctggcacaggttgggttctctgaagatgaagcaaaagcgaggtggatcccgacat 1421
Db 1080 CCTCTGGCACAGGTGTGTGTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1139
Qy 1422 ccttccctgagccctggacagctcctgacgagacgagacccctgagcccaactatcag 1481
Db 1140 CTTTCTCGACCCCTGGACAGTCTCTGACGAGAGCGGAAGACCCCTGAGCCCAACTATGCGAG 1199
Qy 1482 gtgccccctgcctccagggctgtagggtgtgagagaaagggcagggcgaggtccaggg 1541
Db 1200 GTGCCCCCTGCCCCCGAGGCTGTAGGGGTGTGGGAGAAAGGGCGCAGCGGCTCAGGG 1259
Qy 1542 atattgagtgactcttggagctgggctgggtgtgctggcttggcagaaagtcagggct 1601
Db 1260 ATATTGAGTGACTCTTGTGGAGTCTGGGCTGCTGTGGCTTGGCAGTGGCAGTGGGCTTGG 1319
Qy 1602 aagatctcagctcgtgcttggggccctggcaggtgtgtgagccttggctgtctgaca 1661
Db 1320 AAGATCTCATCTCTGTGGCTTGGGGCCCCCTGGCAGGTGTGTGTGTGTGTGTGTGTGTGTGT 1379
Qy 1662 gggaaacaggaggagagacactcggagagtgaggagggcaggtgtgtctgtgagat 1721
Db 1380 GGGAA--CAGGAGGAGGACACAGCTCGGGGAGAGTGGGAGCGCCAGTGTGTGTGTGTGTGTGT 1438
Qy 1722 atgtggccaggttcaagtgggaagctgaaggtatgagcagaccttaggctcaggaaggagg 1781
|||||

Db 1439 ATGTGGCCGACAGGTTTCACTGGGAAGCTGAAGGATGACGACACCTTAGGCTCAGGAAGAGGG 1498
Qy 1782 ctgcttgaagtgggggca tcatcactgaccagaagggaaaactggcagtgccagggt 1841
Db 1499 CTGCTGGAAGTGGGGGATCATCATCTCACCAGAAAGGAAACTGGCAGTGGCAGGGCT 1558
Qy 1842 ggaatggggccctgcattgagcttgaaaaaaactataatagaattggttaccatttttatttt 1901
Db 1559 GGATGGGGCCCTGCATGTGAGGTGTCAAAAACATATAATAGAAATGGTTTACCAATTTTCAATTT 1618
Qy 1902 attatttatttatttatttatttatttatttatttatttatttatttatttatttatttatttattt 1961
Db 1619 ATT 1678
Qy 1962 agtgcgggtggtgtctatctcagctcactgcaacctgtgctcccagagatcaagtattctc 2021
Db 1679 AGTGGGTGGTGCTATCTCAGCTCAGTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACT 1738
Qy 2022 cagctcagctccccaggttagctgggattacaagcatgcaaccacatgctcgggataatt 2081
Db 1739 CAGCCTCAGCCTCCCGAGTAGTGGGATTACAAGCATGCACACCATGCCTTGGATAAAT 1798
Qy 2082 ttgtatttttagttgagacgggtttcaccaaggttggcagactggtctcgaactcttg 2141
Db 1799 TGTGTATTATTTAGTTGAGACGGGGTTTACCAGGTTGGCCAGACTGGTCTCGAACTTCTG 1858
Qy 2142 acctcaggtgatctgcctgcctcggcctcccaaatgctggaattacagatgtaggcac 2201
Db 1859 ACCTCAGGTGATCTGGCTGGCTTGGCTTCCCAAGTGGTGGNAATTTACAGATGTGAGGCAC 1918
Qy 2202 tgcctcggcctgggttaccacattttaaaatggagtgatttaccaccttttatgtygatt 2261
Db 1919 TGTCCCTGGCTGGTTACCCACATTTAAATGGAGTGATTTACCCCTTTTATGTGGATT 1978
Qy 2262 tacagcttggttt 2321
Db 1979 TACAGCT--TTTTTTTTTTTTTTTTTTTTTTTGTAGACAAAGTCTGGTCTGTACACCCAGGCTGGAGT 2036
Qy 2322 gcagtaatgcaatctcagctcactgcaaccttagcctctggtgttcaagcaattctcctg 2381
Db 2037 GCAGTAATGCAATCTCAGCTCAGTCAACCTTAGCCTCTCTGGGGTCAAGCAATTTCTCTG 2096
Qy 2382 cctcagccacactgagtagctggggttacaggcatgcaaccacacagccagggcctaattttt 2441
Db 2097 CCTCAGCCACCTGAGTAG-CTGGGATTACAGGCATGACACACACAGCCAGGCTAATTTTT 2155
Qy 2442 tgtatttttagtagagatgggggtttcggcatgttggcagggctggtctcgaactcctgac 2501
Db 2156 TGTATTATTTAGTAGAGATGGGGTTTTCGCCATGTTGGCCAGGCTGGTCTCGAACTCTCTGAC 2215
Qy 2502 ctcaggtgacccgccccttggcctcccaagtgtctaggtatcaggtgaggaaccaact 2561
Db 2216 CTCAGGTGATTCGCCCGCTTGGCCCTCCCAAGTGCTAGGATTACAGGTGGGAACCACT 2275
Qy 2562 tggccagcctgtggtctatctgttttaaacactgggaaggcctcagccccagggccagact 2621
Db 2276 CGCCAGCCTGTGGGCTATCGTTTAAACACTTGGGAAGGCTGACAGCCCGCAGGCCACAGT 2335
Qy 2622 tagctgagctgagcagttcccagtgccaggtagacgagatgctccacccactactcatg 2681
Db 2336 TAGTGCAGCTGAGCAGTTCCAGTGGCAGGTAGACGATGCTCCACCCACCTACTCTATG 2395
Qy 2682 gctgatctctgtcatagtgagtgcttggcagaccttcatcgttatgggatctcgtg 2741
Db 2396 GCTGATCTCTTGTCTATAGTGAAGTGTCTGGACAGACCTTCATCGTTATGGGATCTCTG 2455
Qy 2742 cccagagtggttgcaatgaaatgggagtggaacaagctcactcgggttagggggcagag 2801
Db 2456 CCCCAGAGTGGGTGGCAATGAATGGGAGTGGACAAGCTCACCTGGGTGTAGGGGCGCAGAG 2515
Qy 2802 ggcggaagtcagaggtgtacccccagagtggtggcagcagagagcttcccagggagatctg 2861
Db 2516 GGCCGAGTCCAGAGTGTACCCCGAGAGTGGGTGGCCAGAGGAGTTCAGGAGGATCTG 2575

```
QY 2862 ggatgagcagagggtgagggaggagagaccagagagggggaactgtggccctgggt 2921
Db 2576 GGATGACAGAGGGTGGAGGAGAGAGACCAGAGAGGGGGAACCTGTGGCCCTGGGT 2635
QY 2922 gggctgagtgctggaggagccagcgccagagagagagagagatggatgggtggcga 2981
Db 2636 GGGTCTGGAGTGGCTGGAGGAGAGCCAGAGCCGAGAGAGAGAGATGGGATGGGTGGGA 2695
QY 2982 gcccagagctgggcccagacctcacactgtgtctgtgtgcccctgcccgtggaccaggtgagcc 3041
Db 2696 GCCCCAGGCTGGGCCGACCTCACACTGTGCTGTGTGCCCTGCCGTGGACCAAGGTGAGCC 2755
QY 3042 atactcgcattcaaggcctacactgtgtgaggggagagagaggtgtccctgctcgaagg 3101
Db 2736 ATACGTCGCCATCAAGGCTACACTGTGTGTGAGGGGACGAGGTGCTCCCTGCTCGAGGG 2815
QY 3102 tgaactgttgaggtaattcaaaactcctggagcgctgggtgggtcatcaggttaggagg 3161
Db 2816 TGAAGCTGTGAGGTAAATTCACAAGCTCCTTGACGGCTGTGTGGTTCATCAGTAGGAGGG 2875
QY 3162 cccctctcattcagagacccattcagtcagcagcccccagagagcgggtgttttaggga 3221
Db 2876 CCCCCTCTCCATCCAGAGACCCATCTGAGTCAGCCCCAGCCAGCGCGTGTTAGGGA 2935
QY 3222 tctgggtgactgtccctggagactctggtaagccactgccctctctggtcttagttt 3281
Db 2936 TCTGGGTGACTTGTCCCTGGGACTCTGGGTAAAGCCACTGCCCTCTCTGGGCTTAGTTT 2995
QY 3282 ccatctcagtagcaggaggatgatgcccaccccttgcctgtcttctgtgggagatccaatg 3341
Db 2996 CCATCTCAGTAGCAGGAGGATGAGCCACCCTTGCTGTCTGTGGGGATCCAAATGTC 3055
QY 3342 cttgtccaaagtgggtgcaatt 3361
Db 3056 CTTGTCCAAGTGGGTGCATT 3075

RESULT 9
LOCUS HS47PHOX02
DEFINITION Homo sapiens p47-phox (NCF1) gene, exons 6-8.
ACCESSION U57834
VERSION U57834.1 GI:2754710
KEYWORDS
SEGMENT
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 3349)
AUTHORS Gorlach,A., Lee,P.L., Roesler,J., Hopkins,P.J., Christensen,B.,
Green,E.D., Chanock,S.J. and Curnutte,J.T.
TITLE A p47-phox pseudogene carries the most common mutation causing
p47-phox- deficient chronic granulomatous disease
J. Clin. Invest. 100 (8), 1907-1918 (1997)
97474758
JOURNAL 2 (bases 1 to 3349)
MEDLINE Chanock,S.J., Roesler,J., Hopkins,P., Lee,P., Bassett,D.T.,
AUTHORS Christensen,B., Curnutte,J.T. and Gorlach,A.
TITLE Genomic Structure and the Identification of Multiple Polymorphisms
in the p47-phox Gene
Unpublished
JOURNAL 3 (bases 1 to 3349)
REFERENCE Gorlach,A., Roesler,J., Christensen,B., Chanock,S.J. and
AUTHORS Curnutte,J.T.
TITLE Direct Submission
JOURNAL Submitted (09-MAY-1996) P. Lee, Molecular and Experimental
Medicine, Scripps Research Institute, 10550 North Torrey Pines Rd,
La Jolla, CA 92037, USA
FEATURES
source
1..3349
/organism="Homo sapiens"
```

```
/db_xref="taxon:9606"
/chromosome="7"
/clone="PI clone 42"
/map="7q11.23"
507..629
/gene="NCF1"
/number=6
1092..1199
/gene="NCF1"
/number=7
2748..2865
/gene="NCF1"
/number=8
BASE COUNT 714 a 860 c 1002 g 773 t
ORIGIN
```

```
Query Match 88.8% Score 2984.4; DB 9; Length 3349;
Best Local Similarity 99.1% Pred. No. 0;
Matches 3053; Conservative 0; Mismatches 21; Indels 6; Gaps 5;
```

```
QY 282 aaaaaaaaaaaaaaagtgtgggaaaaagccaggtgagtggtccacgcctgtatcc 341
Db 1 AAAAAAAAAAAAAAAAAAGTTGGGAAAAAGCCAGGTGCAGTGGCTCCACGGCTGTATCC 60
QY 342 caacactttaagagctgaggtgggagaatcctttgagccagagagttcagaccagct 401
Db 61 CAACACTTTAAGAGCTCAGGTGGGAGAAATCTTTGAGCCCGAGAGTTCGAGACAGCT 120
QY 402 gggcatgtcccaagaccttctttacaaaaaattagccgggtgtgtggtacatctct 461
Db 121 GGGCATTTGCCAAGACCTTCTTTTACAGAAATTAGCCGGGTGTGTGGTCATACGTCT 180
QY 462 gtgtccagctattcggagagctgagccagagagattgcttgagcctagagctaggt 521
Db 181 ATGGTCCCAGCTATTTCGGGAGGCTGAGCGAGGAGATTGCTTGAGCCTTAGGAGTCC 240
QY 522 ctgtagtgcgtgtatcacgtcactgtactctgacctgggcaacacagagcaagactgt 581
Db 241 CTGTAGTCAGCTGTGATCACGTCACGTCCTACTGTACTAGCTGGGCAACAGAGAACTGT 300
QY 582 ctcaaaaaaataaagtgtggaaaggctcactaactcactcagatgagagacaaaga 641
Db 301 CTCCTCAAGAAATAAAGTTGGGAAAGGCTCACTAACTTCATCAGATGAGAAAGGA 360
QY 642 catgttgaagtgtgagccgaagcctgagaaagcctatgcgccagagaaatgcaggga 701
Db 361 CATGTTTGAAGTGTGAGGCCGCAAGCCTGGAGAACCTATGCGCCCGAGAAATGCAAGGA 420
QY 702 gcagagactcaagatgcagcgccctgttctgagggccagagatggccctgcaatgccac 761
Db 421 GCAGAGACTCAAGATGCCAGCGCCTGTTCTGAGGCCCGAGATGGGCCCTGCAATGCCAC 480
QY 762 taacctgcctccctcttgcaccagacatcacccgcccccatcctctgagagactagc 821
Db 481 TCACCTGCCTCCCTCTTGTCCCGCAGACATCACCGGCCCCCATCATCTCTGCAGAGTACC 540
QY 822 cgcattgccactacgagagacctcggtctcgagatggctctgtccacggggagcgt 881
Db 541 GCCATTGCCAACTACGAGAGAGACCTCGGGCTCCGAGATGGCTCTGTCCACGGGGAGCT 600
QY 882 ggtgaggtcgtggagaagagcgagcggttcagacctccacaccttaagggtgctctcc 941
Db 601 GGTGAGGTCTGAGAGAGAGCGAGCGGTCAGACCTCCACACCTTACGGGCTCTCTCC 660
QY 942 cctggtctcaggaaacccacagccccccttcgcaaggctcagcagcctggcccc 1001
Db 661 CCTGTGTCTAGGAACCCACAGCCACCAAGCCCTCTGCCAA-GCTCAGGAGCTGGCCCC 719
QY 1002 tgggagagactcagctctgttagggccctaaatctccccacactgtggtgcctt 1061
Db 720 TGGGAGGACTCCGGCTCTGTAGGGGGCCCTAAATGTCTCTCCACACTGTGTGGTGCCTT 779
```


Db 2935 TCTGGGCTGACTTGCTCCCTGGGACTCTGGGTAAGCCACTGCCCTCTCTCTGGCTTAGTTT 2994
Qy 3282 ccattctcagtagcaggaggagatgagccacccttgctctgtctgtgaggatccaatgtc 3341
Db 2995 CCATCTCAGTAGCAGGAGGAATGAGCCACCCTTGCCCTGCTGTGTGGGGATCCAAATGTC 3054
Qy 3342 cttgtccaagtgggtgcatt 3361
Db 3055 CTTGTCCAAGTGGGTGCATT 3074

RESULT 10
LOCUS HS47P41S03 3200 bp DNA PRI 07-JAN-1998
DEFINITION Homo sapiens p47-phox pseudogene, clone P41, exons 6-8.
ACCESSION U61243
VERSION U61243.1 GI:2754727
KEYWORDS
SEGMENT
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS 1 (bases 1 to 3200)
Gorlach,A., Lee,P.L., Roesler,J., Hopkins,P.J., Christensen,B.,
Green,E.D., Chanock,S.J. and Curnutte,J.T.
TITLE A p47-phox pseudogene carries the most common mutation causing
p47-phox- deficient chronic granulomatous disease
J. Clin. Invest. 100 (8), 1907-1918 (1997)
97474758
REFERENCE 2 (bases 1 to 3200)
Chanock,S.J., Roesler,J., Hopkins,P.J., Lee,P.L., Bassett,D.T.,
Christensen,B., Curnutte,J.T. and Gorlach,A.
TITLE Characterization of the genomic structure of the p47-phox gene
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 3200)
Gorlach,A., Lee,P.L., Roesler,J., Christensen,B., Chanock,S.J. and
Curnutte,J.T.
TITLE Direct Submission
JOURNAL Submitted (19-JUN-1996) Molecular and Experimental Medicine, The
Scripps Research Institute, 10550 N. Torrey Pines Rd., CAL 1, La
Jolla, CA 92037, USA
FEATURES
source
1. 3200
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="P41"
/map="7q11.23"
/note="Human genomic p1 clone P41 is a p47-phox pseudogene
which is defined by the presence of a GT deletion at the
beginning of exon 2."
359..480
exon /pseudo
/number=6
944..1051
exon /pseudo
/number=7
2600..2718
exon /pseudo
/number=8
BASE COUNT 662 a 823 c 971 g 744 t
ORIGIN

Query Match 84.7%; Score 2845.8; DB 9; Length 3200;
Best Local Similarity 99.1%; Pred. No. 0;
Matches 2904; Conservative 0; Mismatches 22; Indels 5; Gaps 4;

Qy 431 aaaaattagccgggtgtggtggcatacgtctgtgtccacgctatttcgggaggtgagcc 490
Db 2 AGAAATTAGCCGGGTGTGTGGTCATACGCTGTGTGGTCCACGCTATTTCGGGGAGGCTGAGGC 61

Db	1141		TGGTGTCTGCTTGGCAGAAAAAGTCAGGGCTAAGATCFATCGTCTTGCCTTGGGGGCC	1200
Qy	1631		ctggcagggttgatgaccccttgctgtgcacagggaacacaggagagagagacagactcg	1690
Db	1201		CTGGCAGGTTGTGATGCCCTTGCTGTGCACAGGNA-CAGAGAGAGAGAGACAGCACTCG	1259
Qy	1691		ggagagtgaggagggcagtggtctctgtggatatgttgggccagggttcagtgggaagctgaag	1750
Db	1260		GGAGAGTGGGAGCCAGTGGTCTGTGTGGATATGTGGCCAGGTTTCAGTGGGAAGCTCAAG	1319
Qy	1751		gatgacgagaccttaggtctcaaggaaggagggtcgcctggaaagtgggggcatcatcactga	1810
Db	1320		GATGACGACACCTTAGGCTCAGGAAGGAGGGCTGCTTGAAGTGGGGCATCATCACTGA	1379
Qy	1811		csagaagaaggaaaaactggcagtgccagggctggatgggggacctgcatgagcttgaaaaaa	1870
Db	1380		CCAGAAAGGGAACACTGGCAGTGCACAGGCTCGATGGGGCTCGATTCAGTTCAGTTGAAAAA	1439
Qy	1871		actataatagaattgggtaccattttattattattattattattattattattattttttt	1930
Db	1440		ACTATAATAGAAATTGGTTACCATTTCATTTTATTTATTTATTTATTTATTTTCTTTT	1499
Qy	1931		agatagagctcaactccttgcttaaggctggagtgcggtgggtggtgtatctcagctcaactgc	1990
Db	1500		AGATAGAGTCTCACTCCCTTGTCTAAGGCTGGAGTGCGGTGTGTATCTCAGCTCACTGC	1559
Qy	1991		aacctgtcctccacagatcaagatgattctccagctcagcctcccaggtagctcaggat	2050
Db	1560		AACCTCTGCTCCAGGATCAAGTGAATTCCTCAGGCTCAGGCTCCCCAGAGTAGCTGGAT	1619
Qy	2051		tacaagcatgcaccaccatgcctggataatttttgatttttagctgagacggggttca	2110
Db	1620		TACAAGCATGCACACCATGCTCGATAAATTGTGTATTTTATTTTATTTTATTTTCTTCA	1679
Qy	2111		caaggtggccagactggtctggaactctgaaactctgaactcagtgatctgctgcctgcctc	2170
Db	1680		CCAGGTTGGCCAGACTGGTCTCGAATTCCTGACCTCAGGTCAGTGTGCTGCCCTTGGGCTC	1739
Qy	2171		ccaaagtctgggaattcacagatgtagcactgtccctggcctgggttacccacatttaa	2230
Db	1740		CCAAAGTGGTGGAAATTCAGATGTCAGGCACTGTCCCTGGGCTCGGTACCCACATTTTAA	1799
Qy	2231		aatggagtgatttcacccctttatggtgatttacagctgtgttttttttttttttttttgaga	2290
Db	1800		AATGGAGTGATTTACCCCTTTATGTGGATTTTACAGCT-TTTTTTTTTTTTTTTTTTTCAGA	1857
Qy	2291		caaagtctggctctgtcacccaggctggagtgagtaagtgaactcagctcaactgaac	2350
Db	1858		CAAAGTCTGGTTCGTGTCACCCAGGCTGGAGTGCAGTAATGCAATCTCAGCTCACTGCAAC	1917
Qy	2351		cttagcctcctgggttcagaactctcctgctcagccactgagtagctcaggtgggttac	2410
Db	1918		CTTAGGCTCCTGGTTCAGCAATTCCTCGCTCAGCCACTCAGTAG-CTGGGATTAC	1976
Qy	2411		aggcatgcaccaccagccaggctaaattttttgtatttttagtagagatggggttcgcc	2470
Db	1977		AGGCATGCACACACGCCAGGCTAAATTTTGTATTTTATTTTATTTTATTTTATTTTCTGCC	2036
Qy	2471		atgttggccaggtggtctcgaactcctgacctcagtgatccgcgcgcgctttggcctccc	2530
Db	2037		ATGTTGGCCAGGCTGTCTCGAATCTCTGACCTCAGGTGATCCGCCGCCCTTGGGCTTCCC	2096
Qy	2531		aaagtctcaggtatcacagtggaaaccaccttgcccagccttggtcattcgttttaaacac	2590
Db	2097		AAAGTCTTAGGATTACAGGTGGGAACCACTCGCCCAAGCCCTGTGGCTATCGTTTAAACAC	2156
Qy	2591		tgggaagcctgcagccccccagccgacagtttagctgcagctgagcagttccagtgcca	2650
Db	2157		TGGGAAGGCTTGACGCCCCAGGCCGACAGTTAGCTGCAGCTGAGCAGTCCCAGTCCCA	2216
Qy	2651		ggtagacgtagtctccaccaccactactcatggctgatctctgttcacatgtaagagtctg	2710

Db	2217	GGTAGACGGATGCTCCACCCACCTACTCATGCTGATCTCTTGTTCATAGTGAAGTGTCTG	2276
Qy	2711	gacagacctcatcgcttatggagatctctgggtcccccagagtggttgccaatgaatggagct	2770
Db	2277	GACAGACCTTCATCGCTTATGGGATCTCTGGTCCCCAGAGTGGTGCCAATGAATGGAGT	2336
Qy	2771	ggacaagtctacctgggtgtatgggggacagagggccgaagtcacagagtgatcccccaagct	2830
Db	2337	GGACAAGCTCACCTGGGTGTAGGGGCGAGAGGGCCGAATCCACAGAGTGTATCCCCAGAGT	2396
Qy	2831	gggtgcccagaggaagcttgccagaggagatctggatggagcagagaggttgagaggagaga	2890
Db	2397	GGGTGCCAGCAGGAGCTTGCCGAGGGATCTGGGATGGAGCAGAGGGTGGAGGGAGGAGA	2456
Qy	2891	cccgaagagggggaacttggtggccctgggtgggtcttgagtgacctggaggaagccaggt	2950
Db	2457	CCCAAGAAGGGGGAACCTGTGGGCCCTGGTGGGTCTGGAGTGCCTGGAGGAAGCCCGAG	2516
Qy	2951	cgcagagggagagaagatgggatgggtggcagagccccagagctgggcccagacctcacactgtg	3010
Db	2517	GCGAGAGAGGAGAAGATGGGATGGGTGGCGAGCCCGAGCTGGGCCGACCTCACACTGTG	2576
Qy	3011	ctctgtgcccttcgctgacagcagtgagccatagccatagctccacatcaagccctacactgtg	3070
Db	2577	CTCTGTGCCCTTCCGCTGACAGGTGAGCCATAGCTGCCCATCAAGGCCCTACACTGCTG	2636
Qy	3071	tggagggggacagaggtgtccctgctcgaggggtgaagctgttgaggttaattcaagctcc	3130
Db	2637	TGGAGGGGACGAGGTGTCCCTGCTCGAGGGTGAAGCTGTTGAGGTAAATTCACAGCTCC	2696
Qy	3131	tggacgggtgtgggtgcatcagtgtagggaggggccctctctcatccagagcaccatctgag	3190
Db	2697	TGGACGGCTGTGGTGTATCAGTATGAGAGGGGCCCTCTCCATCCAGAGCACCACATCTGAG	2756
Qy	3191	tcagccccagccagagcaggggtgttagggatctgggtgactgtccctggagctctg	3250
Db	2757	TCAGCCCCAGCCAGAGCGGGGTATTAGGGATCTGGGGTGACTGTGCCCTGGGACTCTGG	2816
Qy	3251	gtaagccactgcccctctctgggtctagtttccactctcagtagcagggagggatgagccc	3310
Db	2817	GTAAGCCACTGCCCTCTCTGGCTTAGTTCCATCTCAGTAGCAGGAGGGATGAGCCC	2876
Qy	3311	accttgcctgtcttbtggggatcccaatgtctcttccaaatggtgtgcat	3361
Db	2877	ACCCCTTGCTGTCTGTGGGATCCAATGTCTCTGTGCTCAAGTGGGTGCATT	2927
RESULT	11		
HS47L14S02			
LOCUS	2899 bp	DNA	PRI
DEFINITION	Homo sapiens p47-phox (NCF1)	pseudogene, clone L14	07-JAN-1998
ACCESSION	U61225		
VERSION	U61225.1	GI:2754719	
KEYWORDS			
SEGMENT	2 of 2		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostom		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
TITLE	1 (bases 1 to 2899)		
JOURNAL	Gorlach,A., Lee,P.L., Roessler,J., Hopkins,P.J., Christensen,B.,		
MEDLINE	Green,E.D., Chanock,S.J. and Curnutte,J.T.		
REFERENCE	A p47-phox pseudogene carries the most common mutation causing		
AUTHORS	p47-phox- deficient chronic granulomatous disease		
TITLE	J. Clin. Invest. 100 (8), 1907-1918 (1997)		
JOURNAL	97474758		
MEDLINE	2 (bases 1 to 2899)		
REFERENCE	Chanock,S.J., Roessler,J., Hopkins,P.J., Lee,P.L., Bassett,D.T.,		
AUTHORS	Christensen,B., Curnutte,J.T. and Gorlach,A.		
TITLE	Characterization of the genomic structure of the p47-phox gene		
JOURNAL	Unpublished		
REFERENCE	3 (bases 1 to 2899)		
AUTHORS	Gorlach,A., Lee,P.L., Roessler,J., Chanock,S.J. and Curnutte,J.T.		

TITLE Direct Submission
JOURNAL Submitted (18-JUN-1996) Molecular and Experimental Medicine, The Scripps Research Institute, 10550 N. Torrey Pines Rd., CAL 1, La Jolla, CA 92037, USA
FEATURES Location/Qualifiers
 source 1..2899
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="L14"
 /chromosome="7"
 /map="7q11.23"
 /note="Human genomic EMBL3A bacteriophage clone L14 is a p47-pho pseudogene which is defined by the presence of a 289 bp deletion at the beginning of exon 2"
gene join(U61224.1:<879..8119,1..2866)
 /gene="NCF1"
 /pseudo
exon 507..629
 /gene="NCF1"
 /pseudo
exon 1091..1199
 /gene="NCF1"
 /pseudo
exon 2749..2866
 /gene="NCF1"
 /pseudo
 /number=8
BASE COUNT 618 a 740 c 882 g 659 t
ORIGIN

 Query Match 84.0%; Score 2823.6; DB 9; Length 2899;
 Best Local Similarity 99.2%; Pred. NO. 0;
 Matches 2880; Conservative 0; Mismatches 19; Indels 5; Gaps 4

 Qy 282 aaaaaaaaaaaagtgtggaaaggccaggtgcagtggtccacgcctgaatcc 341
 |||||
 Db 1 AAAAAAAAAAAAAAAAAAGTTGGGAAAGCCAGGTGGCTCCAGCGCTGTAATCC 60

 Qy 342 caacactttaagggtgcaggtggagaaatcctttgagcccgaggttcgagaccagcct 401
 |||||
 Db 61 CAACACATTTAAGAGGCTGAGTGGGAGAACTCTTTAGCCCGCAGAGTTCGAGACGACGCT 120

 Qy 402 ggcattgtccaaagacctgtctttacaaaaattagccgggtgtggtggcatacgtct 461
 |||||
 Db 121 GGGCATGTGCTCCCAAGACCTTGCTCTTACCAGAAATTAGCCGGGTGTGTGGCATACGCTCT 180

 Qy 462 gtggtccagctattcgggaggtcaggcaggagagattgcttgagcctaggagctaggg 521
 |||||
 Db 181 GTGGTCCCAACTATTCCGGAGGCTGAGGCAGGACAGATTGCTTGAGCCTAGGAGTCTAGGG 240

 Qy 522 ctgtagtgaactgtgatacgtcaactgtacttaagcctgggcaacagagaagactctgt 581
 |||||
 Db 241 CTGTAGTGTGAGCTGTGATCAGCTACGTGTACTCTAACCCTGGGCAACAGAGCAAGACTCTGT 300

 Qy 582 ctccaaaaaagaaaaTaaagtgtggaaaggtccactcaacttcacagatgagaacaaaga 641
 |||||
 Db 301 CTCCTCAAAAGAAATAAAGTTGGGAAAGGCTCACTAACTTCATCAGATGAGAAACAAGGA 360

 Qy 642 catgtttgaagtgtgaggccgaagcctggagaaacgctatgcgccagagaaatgcagggca 701
 |||||
 Db 361 CATGTTTGAAGTGTGAGGCGGAAGCGTGGAGAACGCTATGCGCCAGGAAATGCAGGGCA 420

 Qy 702 qcagaactcaagatccagcgccctgttctggagcccgagatggccctgcaatgccac 761
 |||||
 Db 421 GCAGAGACTCAAGATGCCAGCGCCTGTTCTGGAGGCCCAGATGGCCCTGCAATGCCAC 480

 Qy 762 tcaccttgccctccctcttgccccagacataccgggcccccatcatcctgcagagctaccg 821
 |||||
 Db 481 TCACCTCTGCCCTCTTGTGCCCCAGACATACCGGGCCCCATCATCTTGCAGAGCTACCG 540

Db	1619	ATATATTTATTTATTTATTTATTTTACATTTTTCAGATAGAGTCTCACCTCTTGTCTAAGGCTGG	1678
Qy	1962	agtgcggtggtcgtatctcagctcactgcaaacctctgctcccaaggataccaagtatcttc	2021
Db	1679	AGTCGGTGTGTATCTCAGCTCACTGCAACCTCTGCTCCCAAGGATCAAGTGATTTCTC	1738
Qy	2022	cagcctcagcctcccaggtagctgggatcacagcatgcaccaccatgcctcgtgataatt	2081
Db	1739	CAGCCTCAGCCTCCCAAGTACGTGGATTTACAAAGCATGTCCACCACTGCGTGGATAA	1798
Qy	2082	tttgtatttttagtgagcggggtttcaccaggttgccagagctgctcgaactctg	2141
Db	1799	TTTGTATTTTAGTTCAGACGGGGTTTACCAGGTTGCCAGACTGGTCTCGAACTTCTG	1858
Qy	2142	acctcaggtgatatgctcgtcctcgccctcccaaatgctggaattacagatgtgagccac	2201
Db	1859	ACCTCAGGTGATCTGCTGCTTGCCCTCCCAAAAGTCTGGAATTCAGATGTGAGCCAC	1918
Qy	2202	tgctcctggcctgggtttaccacaattttaaaatgaggtgatttcaaccttttatctgga	2261
Db	1919	TGTCCCTGGCCTGGTTTACCACATTTTAAATTTGAGATGATTTTACCTCTTATCTGAT	1978
Qy	2262	tacagctgttt	2321
Db	1979	TACAGCT- -TTTTTTTTTTTTTTTGTAGACAAAGTCTGGTCTGTGTACCCAGGCTGAG	2036
Qy	2322	gcagtaatgcaactcagctcactgcaaaccttagcctcctgggttccaaagcaattcctcg	2381
Db	2037	GCAGTAATGCAATCTCAGCTCACTGCAACCTTAGCCTCTGGGTTCAAGCAATTTCTCTG	2096
Qy	2382	cctcagccactgagtagcctgggtttacaggcatgcaccaccacgcagagcttaattttt	2441
Db	2097	CCTCAGCCACTGAGTAG-CTGGGATTTACAGGCATGTCAACCACTGCTCGAACCTCTG	2155
Qy	2442	tgtatttttagtagaatgggttttcgcaatgtggccaggtggtctcgaaactcctgac	2501
Db	2156	TGTATTTTGTAGATGGGGTTTGGCCGATGTTGGCCAGGCTGGTCTCGAACCTCTGAC	2215
Qy	2502	ctcaggtgataccgcccgccttgccctcccaagtgctaggattacaggttggaaccacct	2561
Db	2216	CTCAGGTGATCCGCCGCCCTTGCCCTCCCAAGTCTAGGATTTACAGGTGGGAACCACT	2275
Qy	2562	tgcccagccttggtctatcgtttaaacactgggaaggcctgcagccccagggcgacagt	2621
Db	2276	CGCCCAAGCCTGTGGCTTATCGTTTAAACACTGGGAAGGCTGCAGCCCCAGCGGACAG	2335
Qy	2622	tagctcagctaaagagttcccgatgcccaggtagacggatgctccaccaccctactcatg	2681
Db	2336	TAGCTGCAGCTGAGCAGTTTCCCAAGTGGCAGGTAGACGGATGCTCCACCCACTACTCATG	2395
Qy	2682	gctgattctctgtcatagtgaaagtctctggacagaccttcagtattggatctctggt	2741
Db	2396	GCTGATCTCTGTTCATAGTAGTGTCTGGACAGACCTTTCATCTGTTATGGGATCTCTGG	2455
Qy	2742	ccccagagtgggtgggcaatgaatggagtggaagctcactgggtgtagggggcagag	2801
Db	2456	CCCCAGAGTGGTGGCAATGAATGGAGTGGACAAGCTCACTGGGTGTAGGGGGCAGAG	2515
Qy	2802	ggccgaagtcagagtgatacccccaagatgggtgccacagagagcttgcagagggatctg	2861
Db	2516	GGCCGAAGTCCAGAGTGTACCCCCAGAGTGGGTGCCACAGAGAGCTTCCCGAGGGATCTG	2575
Qy	2862	ggatggacagagaggtgagggagagaccagaaagggggaactgtgggccctgggt	2921
Db	2576	GGATGGACAGAGGAGTGGAGGGAGGAGACCCAGAAAGAGGGGAACCTGTGGGCCCTGGGT	2635
Qy	2922	ggatctggagtgcttgaggaagcccgagagaggaagaatgggaatgggtggcgga	2981
Db	2636	GGCTCTGGAGTGCCTGGAGGAAGCCAGCGCGCAGAGAGGAAAGATGGGATGGGTGGCGA	2695
Qy	2982	gccccaggtggccagacctcacactgtctctgtgcacctgcccgtgggacaggtgagcc	3041

Db 2696 GCCCAGCGCTGGCGGACCTCACACTGTGCTCTGTGCCCTGGCTGGACGACGAGGAGGCC 3165
 Qy 3042 atacgtgcacatcaagcgcctacactcgtgtgaggggagcagaggtgtccctgtgctgaggg 3101
 Db 2756 ATACGTGCCCATCAAGGCGCTACACTGCTGTGGAGGGGACGAGGTGTCCCTGCTCCGAGG 2815
 Qy 3102 tgaagctgttgaggttaattcaacagctcctcctgagcggcgtgtggtgagctcaggtgagggg 3161
 Db 2816 TGAAGCTGTTCAGGTCTCATTTCAACAGCTCCCTGGACGGCTGGTGGCTCATCAGTAGGAGG 2875
 Qy 3162 cccctctccatccagagaccat 3185
 Db 2876 CCCCTCTCCATCCAGAGCACCCAT 2899

RESULT 12
 AC092405/c
 LOCUS
 DEFINITION Papio cynocephalus anubis clone RP41-170F23, WORKING DRAFT
 AC092405
 VERSION AC092405.1 GI:14595779
 KEYWORDS HTG: HTGS_PHASE1; HTGS_DRAFT.
 SOURCE Olive baboon.
 ORGANISM Papio cynocephalus anubis
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
 Cercopithecinae; Papio.
 1 (bases 1 to 184558)
 Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B., Blakesley,R.W.,
 Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S., Dietrich,N.L.,
 Granite,S., Guan,X., Gupta,J., Ho,S.-L., Idol,J.R., Karlins,E.,
 Lee-lin,S.-Q., Legaspi,R., Lim,M., Maduro,Q.L., Maduro,V.B.,
 Masello,C., Mastriani,S.D., McCloskey,J.C., McDowell,J.,
 Pearson,R., Prasad,A., Shevchenko,Y., Snyder,B., Stantipop,S.,
 Thomas,J.W., Thomas,P.J., Touchman,J.W., Tsurgon,C., Vogt,J.L.,
 Walker,M.A., Wetherby,K.D., Zhang,L.-H. and Green,E.D.
 NISC Comparative Sequencing Initiative
 Unpublished
 2 (bases 1 to 184558)
 Green,E.D.
 Direct Submission
 Submitted (04-JUL-2001) NIH Intramural Sequencing Center, 8717
 Government Circle, Gaithersburg, MD 20877, USA
 ----- Genome Center
 Center: NIH Intramural Sequencing Center
 Center code: NISC
 Web site: <http://www.nisc.nih.gov>
 Contact: nisc_mouse@hri.nih.gov
 ----- Project Information
 Center project name: CCX
 Center clone name: 170F23
 ----- Summary Statistics
 Sequencing vector: plasmid; n/a; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.990319
 Consensus quality: 178116 bases at least Q40
 Consensus quality: 179898 bases at least Q30
 Consensus quality: 181141 bases at least Q20
 Insert size: 168000; agarose-
 Insert size: 184058; sum-of-
 Quality coverage: 10.52x in Q20 bases; agarose-
 Quality coverage: 9.60x in Q20 bases; sum-of-

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 6 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 *
 * 1 13828: contig of 13828 bp in length

Db 142926	GAGTACTGCTTTGGAGTCTGAGACTGCTTGGCTGGCGAGAAAAGTCAGGGCTGAGAC	142867
Qy 1607	ctcatcgccctggcctt-gggggcccctggcaggttgatgcctctggcttgcttgacacgga	1665
Db 142866	CTCATCTTCTGCTTGGGGGGCCCTGGCAGGTTGGGATATCCTTGGTCCGGACAGGA	142807
Qy 1666	accgagggagagacagacactcggagagtggggagggccagtggtgtctgtggtatgt	1725
Db 142806	ACCAGGAGGAGGACAAACATGGGGGAGGGTGGGAGTTCAGTTATGTCTATGGGTATGT	142747
Qy 1726	ggccaggttcagttgggaagctgaaggtgagcagacttagctcaggaagagggctgc	1785
Db 142746	GGTCAGGTCCAGTGGGAAACTGAAGGACGAGCAGACTTAGGTTCAGGAAGAGGGCTGC	142687
Qy 1786	ctggaagtgggggcatcatcactgaccagaaaggaaaaactggcagtgccagggctggat	1845
Db 142686	CTGAAGGTGGAGGCGTCATCAGTCAGCCAGAAAGGAAAACTGSCAGTGTGGGCTGGAT	142627
Qy 1846	ggggcctgcattgagct-tgaaaaaaactaataagaattgggttaacctttatttta-t	1903
Db 142626	GGGGCCTGCATTGAGCTCGGAAAAAACTAATATAGAAATGGTTACCATTTATTTTACT	142567
Qy 1904	tatttattttattttact- ----tctttgagatagacttcactccctgtctaaggc	1958
Db 142566	TATTATTTATTTATCTTGTGTGTGTGTGTGAGACAGCGTCTCACTCCCTGCTCAGCG	142507
Qy 1959	tggagtcgggtggctcatctcagctcactgcgaacctctgc-ctcccaggatcaagtgat	2017
Db 142506	TGGAGTGCAGTGGTGATCTCGGTTCAGTGCAATCTCTGCTGTCCGGGTTCAGTGAT	142447
Qy 2018	tctcagcctcagctcccccaggtagctgggattacaagcatgcacacacatgcctggat	2077
Db 142446	TCTCAGCGCTCAGCGCTCTCGA-GTAGTGGGATTACAGGCATGTGCCACACACCCAGCT	142388
Qy 2078	aattttgtatttttagttgagacggggtttccaccaggttggccagactgggtctcgaact	2137
Db 142387	AATTTTGTATTTTAAATAGAGAC-GGGTTTTCACCATTTTGGCGCAGGCTGGTCTCGA	142329
Qy 2138	tctgaactcaggtgatctgctgcctgcgctcccaaaagtctggaattacagatgatgag	2197
Db 142328	TCTGACCTCAGGTAAATCGGCTGCCTCAGTCTCCCAAGGGCTGGGATTTACAGGTATG	142269
Qy 2198	ccactgtccctggcctgggttaccocaaatttaaaaggagtgatttcaaccttttatgtg	2257
Db 142268	CCACTGCCCCCTGGCCTGGTTACCCACATTTTAAATGGCGTGATTTCACTTTTATGTG	142209
Qy 2258	gatttaacag- ----tgttt	2312
Db 142208	GATTTACAGCGCTTTTtt	142149
Qy 2313	ggctggagtcagtaatgcgaactcagctcactgcgaaccttagctccctgggttcaagca	2372
Db 142148	GGCTGGAGTGTAGTAAATGCAAATCTTGGCTCAGTGCAACCTTCACTTCTGGGTTCA	142089
Qy 2373	attctctgcctcagccacctgagtagcctggggttacaggaatgcaccaccacgcagag	2432
Db 142088	ATTCTCCACCTCAGCGCTCCTGAGTAG-CTGGGATTACAGGCGTGCCACACCGCCAG	142030
Qy 2433	ctaatttttgtatttttagtagagatgggggtttcgccatgttggccaggtcgggtctcga	2492
Db 142029	CTAA-TTTTTGTATTTTGTAGATGGGGTTTTCACCATTTTGGAGACGGCTGGTCTTGA	141971
Qy 2493	actcctgaactcaggtgatccggccgcttggcctcccaagtgtcagattacagttgg	2552
Db 141970	ACTCCTGACCTCAGGTGATCCACCCACTCGGGCACCCAAAGTGCGGGGATTACAGCGG	141911
Qy 2553	gaacacacttgcacgcttggctatcggttttaaaactgggaagggcctgcagcccccgag	2612
Db 141910	GAACCACTTGTCCAGCTGTGGCTATTGTTTAAACACTGGGAAGGCTTCAGCCCCCAG	141851
Qy 2613	gccgacagtttagctgcagctgagcagttcccatgcccaggtagacgga-tgcttccacccac	2672
Db 141850	GCCCCAG- ----AGGCCACAGCCAAAGCAGTTCCAGTGGCAGCTGCAGAGGCGCTCACCC	141795

Qy	2673	ctactcagggctgatactcttgcatagttgtaagtgctgacagaccttcatcgttatggg	2732
Db	141794	CVACTCATGGCTGGGTCTCTGTGTCATATTTAGGTGCTGGACAGACCTTCATCGTTTGGGA	141735
Qy	2733	atctctggtccccagagtggtgtggcaatgaatggagtgagacaagctcaccttgggttag	2792
Db	141734	A--TCTGTTCCCGAGTGGGTGGCAATGAATGGGAGTGGACAAGCTACCTGGGTGTAG	141677
Qy	2793	ggggcagagggccgaagctccagagtacccccagagtgsggtgccagcagagagcttgcg	2852
Db	141676	GGGGCAGAGGGCAGAAAGTCCAGAGTGTGCCCCAGAGAGGGTGCCACGACGAGCTTCCGC	141617
Qy	2853	aggg-atctgggatggagcagagggctggaggagagaccaggaagaggggaaactgtg	2911
Db	141616	AGGGAAACCCGGATGGAGTAGAGGATGGGGGAGAGACCCAGAAGA-GGGGCGCTGTG	141558
Qy	2912	ggccctgggtgggtctggagtgccctggagaaagccaggccgacagaggaagatggga	2971
Db	141557	GGCCCTGGGTGGGTCTGGAGTGCCCTGGAGAGAACCCAGGTCGAGAGAGAAGACGGGA	141498
Qy	2972	tgggtggcgagccccagagctggggcgagctcacactgtctctgtgtccctgcggtgac	3031
Db	141497	TGGGTGGCGAGCCCGCGGTGGCCAGATCTCACCCGTTCTGTGTGCCCGCGGTGGAC	141438
Qy	3032	caggtgagccatacgtcgcacataagcctacactgctgtggaggggagcagaggtgtccc	3091
Db	141437	CAGGCGAGCCGTACGTTGCCATCAAGCCGTACGCTGCTGTGGAGGAGGACGAGGTGTCCC	141378
Qy	3092	tgctcgaagggtagagctgttgagtgtaattcacaaagctcctggacggctggtgggtcatca	3151
Db	141377	TGCTCGAGGGCGAAGCTGTTAGGTCTATTACAAGCTCCTGGACGGCTGCTGGGTCTATCA	141318
Qy	3152	ggtagaggggccctctccatccagagaccccaactgagtcagccccagccagacggagg-	3210
Db	141317	GGTAGAGGGGCTCCGCTCCATCCAGGGGCACCCACCGGAGTTCAGCCCCACACGGGTGGGA	141258
Qy	3211	gtggttagggatctgggtgactgtccctgggaactctgggtaagccactgcccctctct	3270
Db	141257	CTGTTTGGGGATCTGGGTGGACTGTTCCTTGGGACTCTGGGTGAAGGCACATGCCTCTCTCT	141198
Qy	3271	gggcttagtttccactcctagtagcagggagggatgagccacccttgcctgtcttgtggg	3330
Db	141197	GGGCTCAGTTTCCATCTCAGTAGCAGGGAGGGTGAGCCACCCACCTGGCCCTGTCTTGTGGG	141138
Qy	3331	gatccaatgtcctctgtcccaagtgggtgcac	3360
Db	141137	GATCCAAATGCTCTGTGTCGAAGTGCATGTCAT	141108

RESULT	13
LOCUS	AC027219/c
DEFINITION	Homo sapiens clone RP11-729P19, *** SEQUENCING IN PROGRESS ***, 799 bp unordered pieces.
ACCESSION	AC027219
VERSION	AC027219.3 GI:13184220
KEYWORDS	HTG; HTGS_PHASE1.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 149830)
AUTHORS	Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE	Homo sapiens, clone RP11-729P19 .
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 149830) Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,

Anderson, S., Baldwin, J., Barna, N., Bastien, V., Beda, F., Boguslavsky, L., Boukhgalter, B., Brown, A., Burkett, G., Campopiano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P., DeArellano, K., Dewar, K., Diaz, J.S., Dodge, S., Domingo, M., Doyle, M., Ferreira, P., FitzHugh, W., Gage, D.,

```

Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., LaRoque,K., Lamazares,R., Landers,T., Lehoczy,J.,
Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
Meldrum,J., Meheus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neill,D., Olivari,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tessaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (28-MAR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 2, 2001 this sequence version replaced gi:11610941.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence.submissions@genome.wi.mit.edu
----- Project Information
-----
Center project name: L8346
Center clone name: 729_P_19
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 79 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 1569: contig of 1569 bp in length
* 1570 1669: gap of 100 bp
* 1670 2364: contig of 695 bp in length
* 2365 2464: gap of 100 bp
* 2465 3099: contig of 635 bp in length
* 3100 3199: gap of 100 bp
* 3200 3762: contig of 563 bp in length
* 3763 3862: gap of 100 bp
* 3863 4534: contig of 672 bp in length
* 4535 4634: gap of 100 bp
* 4635 5474: contig of 840 bp in length
* 5475 5574: gap of 100 bp
* 5575 6460: contig of 886 bp in length
* 6461 6560: gap of 100 bp
* 6561 7785: contig of 1225 bp in length
* 7786 7885: gap of 100 bp
* 7886 8735: contig of 850 bp in length
* 8736 8835: gap of 100 bp
* 8836 9864: contig of 1029 bp in length
* 9865 9964: gap of 100 bp
* 9965 10784: contig of 820 bp in length
* 10785 10884: gap of 100 bp
* 10885 11653: contig of 769 bp in length
* 11654 11753: gap of 100 bp
* 11754 12415: contig of 662 bp in length
* 12416 12515: gap of 100 bp
* 12516 13232: contig of 717 bp in length
* 13233 13332: gap of 100 bp
* 13333 14178: contig of 846 bp in length
* 14179 14278: gap of 100 bp
* 14279 15762: contig of 1484 bp in length
* 15763 15862: gap of 100 bp
* 15863 16423: contig of 561 bp in length
* 16424 16523: gap of 100 bp
*
* 16524 17467: contig of 944 bp in length
* 17468 17567: gap of 100 bp
* 17568 18533: contig of 966 bp in length
* 18534 18633: gap of 100 bp
* 18634 19882: contig of 1249 bp in length
* 19883 19982: gap of 100 bp
* 19983 21070: contig of 1088 bp in length
* 21071 21170: gap of 100 bp
* 21171 22152: contig of 982 bp in length
* 22153 22252: gap of 100 bp
* 22253 23303: contig of 1051 bp in length
* 23304 23403: gap of 100 bp
* 23404 24381: contig of 978 bp in length
* 24382 24481: gap of 100 bp
* 24482 26183: contig of 1702 bp in length
* 26184 26283: gap of 100 bp
* 26284 27663: contig of 1380 bp in length
* 27664 27763: gap of 100 bp
* 27764 28566: contig of 803 bp in length
* 28567 28666: gap of 100 bp
* 28667 30309: contig of 1643 bp in length
* 30310 30409: gap of 100 bp
* 30410 31434: contig of 1025 bp in length
* 31435 31534: gap of 100 bp
* 31535 32225: contig of 691 bp in length
* 32226 32325: gap of 100 bp
* 32326 33690: contig of 1365 bp in length
* 33691 33790: gap of 100 bp
* 33791 34434: contig of 644 bp in length
* 34435 34534: gap of 100 bp
* 34535 35776: contig of 1242 bp in length
* 35777 35876: gap of 100 bp
* 35877 37225: contig of 1349 bp in length
* 37226 37325: gap of 100 bp
* 37326 38564: contig of 1239 bp in length
* 38565 38664: gap of 100 bp
* 38665 40341: contig of 1677 bp in length
* 40342 40441: gap of 100 bp
* 40442 41665: contig of 1224 bp in length
* 41666 41765: gap of 100 bp
* 41766 42998: contig of 1233 bp in length
* 42999 43098: gap of 100 bp
* 43099 44123: contig of 1025 bp in length
* 44124 44223: gap of 100 bp
* 44224 46278: contig of 2055 bp in length
* 46279 46378: gap of 100 bp
* 46379 47729: contig of 1351 bp in length
* 47730 47829: gap of 100 bp
* 47830 49626: contig of 1797 bp in length
* 49627 49726: gap of 100 bp
* 49727 51471: contig of 1745 bp in length
* 51472 51571: gap of 100 bp
* 51572 52901: contig of 1330 bp in length
* 52902 53001: gap of 100 bp
* 53002 53821: contig of 820 bp in length
* 53822 53921: gap of 100 bp
* 53922 55355: contig of 1434 bp in length
* 55356 55455: gap of 100 bp
* 55456 57801: contig of 2346 bp in length
* 57802 57901: gap of 100 bp
* 57902 59549: contig of 1648 bp in length
* 59550 59649: gap of 100 bp
* 59650 60555: contig of 906 bp in length
* 60556 60655: gap of 100 bp
* 60656 62781: contig of 2126 bp in length
* 62782 62881: gap of 100 bp
* 62882 64370: contig of 1489 bp in length
* 64371 64470: gap of 100 bp
* 64471 65603: contig of 1133 bp in length
* 65604 65703: gap of 100 bp
* 65704 67667: contig of 1964 bp in length
* 67668 67767: gap of 100 bp
* 67768 69279: contig of 1512 bp in length
*

```

TITLE
JOURNAL

COMMENT

Query Match	51.28;	Score 1719.8;	DB 2;	Length 149830;
Best Local Similarity	94.09;	Pred. No. 0;		
Matches 1748;	Conservative 0;	Mismatches 107;	Indels 4;	Gaps 2;
Qy 931	gggtctctcccttgggtcaggaacccacagcaacagccctcgtccaaggtccagcg	990		
Db 51471	GGGCTCTCCCTGGTGTCTCAGGAACCCACAGCCAAAGCCCCCTGCCAGGTCAGGC	51412		
Qy 991	agcctggccctggaggactccagctctgttagggggccctaaatgtcctccccacactg	1050		
Db 51411	AGCCTGGCCCCCTGGAGGACTCCACAGCTGTGTAGGGGCCCTAAATGTCTCCCAACATG	51352		
Qy 1051	tgggtgcctctctcttagtgcacccctgtgtgctgtgggcatctgtgcattggcag	1110		
Db 51351	TGGGTGCGCTTCTCTTAGTGTGCACCTGTGTGGCTGTGGGCATCTGTGCATGGCAG	51292		
Qy 1111	gccggggcgggcagctctgcgttctctgtcgtgagtgagtgagacgctgtgtcatta	1170		
Db 51291	GCCGGGCGGGCATGTCTGGTGTCTGTGATGGGTATGGACCGCTGTGTTCATTA	51232		
Qy 1171	tgaagtgggtcagagctgtgattctgttgagcatgtgtgcattgcattgacacctat	1230		
Db 51231	TGAAGTGGGCTCAGAGCTGTGATTCTGTGAGCATGTGTGCATGCATGTGACCTCAT	51172		
Qy 1231	tgtccagtggtgaaagtgacatttccaaatctgagcattggacatcaagtgtctctgtg	1290		
Db 51171	TGTCAGCTGTGGTGAAGGTGACATTTCCAAATCTGAGCATTTGGACATCAGTGTGTCTGTG	51112		
Qy 1291	tccctgtctctcaccatccctgatggctgcagggagccgctggggccctgccctcagtc	1350		
Db 51111	TCCCTGTGTCTCACCATCCCTGATGGCTGCAGGGAGCCGCTGGGCCCTGGCCCTCAGTC	51052		
Qy 1351	acattcccacacctctggcacaggttgggttctgtcagatgaagcaaacgagcggtct	1410		
Db 51051	ACATTCCCGCACCTCTGSCACAGGTTGGTGGTTCTGTGCAGATGAAGCAAAAGCAGGCGTG	50992		
Qy 1411	gattccagcatcttctctgaagcccttgacagtcctgtgcagacgacgaagacctgagcc	1470		
Db 50991	GATCCAGCATCTTCTCTCAGGCCCTTGACAGTCCTTACAGAGCGAAGAACCTTGAGCC	50932		

QY 2551 ggaaccacccctgcccagcctgtggtatcggtttaaacacacagggaagcctgcagccccc 2610
 Db 49855 GGGAAACACCTTGGCCAGCCTGTGGCTATCGTTTAAACACTGGGAAGGCTGCAGCCCCC 49796

QY 2611 agccgacagtttagctgcagctgacagctccagttccagtgccaggtagacgagtgctccaccc 2670
 Db 49795 AGCGCGACAGTTAGTGTGAGCTGAGCAGTTCACAGTGCAGGTAGACGATGCTCCACCC 49736

QY 2671 acctactggtgatctctctgtcatagtgaagtgtctgacagacaccttcacgttatg 2730
 Db 49735 ACCTACTCANN 49676

QY 2731 ggaatctgtgtcccccagagtggtggcaatgaatggagtgacagctcacctcaggtg 2789
 Db 49675 NNN 49617

RESULT 14
 HSNCF154
 LOCUS
 DEFINITION Homo sapiens p47-phox (NCF1) pseudogene, clone P38, exons 6 and 7.
 ACCESSION U69642
 VERSION U69642.1 GI:2754733
 KEYWORDS
 SEGMENT 4 of 5
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 918)
 AUTHORS Gorlach,A., Lee,P.L., Roesler,J., Hopkins,P.J., Christensen,B.,
 Green,E.D., Chanock,S.J. and Curnutte,J.T.
 TITLE A p47-phox pseudogene carries the most common mutation causing
 p47-phox- deficient chronic granulomatous disease
 J. Clin. Invest. 100 (8), 1907-1918 (1997)
 JOURNAL 97474758
 MEDLINE
 REFERENCE 2 (bases 1 to 918)
 AUTHORS Chanock,S.J., Roesler,J., Hopkins,P.J., Lee,P.L., Bassett,D.T.,
 Christensen,B., Curnutte,J.T. and Gorlach,A.
 TITLE Characterization of the genomic structure of the p47-phox gene
 Unpublished
 JOURNAL 3 (bases 1 to 918)
 REFERENCE Gorlach,A., Lee,P.L., Roesler,J., Christensen,B., Chanock,S.J. and
 Curnutte,J.T.
 JOURNAL Direct Submission
 TITLE Submitted (06-SEP-1996) Molecular and Experimental Medicine, The
 JOURNAL Scripps Research Institute, 10550 N. Torrey Pines Rd., CAL 1, La
 Jolla, CA 92037, USA
 FEATURES
 source
 Location/Qualifiers
 1..918
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="7"
 /map="q11.23"
 /clone="p38"
 /note="pseudogene is defined by the presence of a GT
 deletion at the beginning of exon 2"
 20..143
 /gene="NCF1"
 /number=6
 /pseudo
 606..713
 /gene="NCF1"
 /number=7
 /pseudo

exon 163 a 261 c 293 g 201 t

exon

exon

BASE COUNT 163 a 261 c 293 g 201 t

ORIGIN

Query Match 26.4%; Score 886.4; DB 9; Length 918;
 Best Local Similarity 99.1%; Pred. No. 8e-217; Gaps 2;
 Matches 912; Conservative 0; Mismatches 6; Indels 2;

QY 768 tgcctcctcttggccccagacatcacccggcccccatcctctgcagcgtaccgcgccc 827
 Db 1 TGCCCTCCCTCTTGGCCCCAGACATCACCGGCCCATCATCTCTGCAGAGTACCGGCCCAT 60

QY 828 tgcgaactacgagaagacctcggtccagatgctctctgtccacgggggagcgttgtgga 887
 Db 61 TGCCAACTACGAGAAGACCTCGGGCTCGGAGATGCTGTGTCCACGGGGAGCGTGGTGA 120

QY 888 ggtcgtgagaagcgcagagcggtcagacctccacaccttacgggggctcctccctcgtg 947
 Db 121 GGTCTGTGAGAAGACGAGAGCGGTTCAGACCTCCACACCTTACGGGGCTCCTTCCCTGGT 180

QY 948 gctcaggaaacacacagcagccccctgccaaaggtcagcagcctggccccctggag 1007
 Db 181 GCTCAGGAACCCACAGCCACAAAGCCCCCTGGCAA -GCTCAGGCAGCCTGGCCCCCTGGGAG 239

QY 1008 gactccagctctgttagggggccctaaatgtctctccacacctgtgggtcgcctctctct 1067
 Db 240 GACTCAGGCTCTGTTAGGGGCCCTAAATGTCTCTCCCACTGTGGGTGCGCTTCTCTCT 299

QY 1068 tagtgtcacccctgtgtgtgtgtggtcatctgtcagtcagcggcgggggcgcatgt 1127
 Db 300 TAGTGTGCACCCTGTGTGGCTGTGGGCATCTGTGCATGGCAGCGCGGGGGGCGCATGT 359

QY 1128 ctgcgtgttctgtctggatgggtatgggacctctgttcaattatgaagtgggctcagagc 1187
 Db 360 CTGCGTGTCTCTCTGGATGGGTATGGGACCGCTCTGTTTATTATGAAGTGGCTCAGAGC 419

QY 1188 tctgattctgtgagcatgtgtcatgtcatgtgacacctatgtccagttgtgtgaag 1247
 Db 420 TGTGATTCTGTGAGCATGTGTGCATGTGCATGTGCATCTATTGTCCAGTGTGTGAAG 479

QY 1248 gtgacatttccaaatctgacgattgacatcagtgctgtctgtccctgtgtcccaacca 1307
 Db 480 GTGACATTTCCAAATCTGAGCAATTTGACATTCAGTGTGTGTGTCTGTGTCTGTCTCCCA 539

QY 1308 tccctgatggtgcaggagccgctgggacctgccccctgcacctcaagtccacacctctg 1367
 Db 540 TCCCTGATGGCTGCAGGAGCGGCTGGGGCCCTGCCCCCTCAGTCACATTCGCCGACCTCTG 599

QY 1368 gcaaggttgggttctctcagatgaaagcaagcaggtggtggtccagacatcttcc 1427
 Db 600 GCACAGGTTGGTGGTCTGT 659

QY 1428 tcgagccctgacagctcctgcagacgaggaagccctgcagcccaactatgcagtgccc 1487
 Db 660 TCGAGCCCTTGACAGTCTCTGACGAGACGGAAGACCCCTGAGGCCCAACTATGAGGTGCC 719

QY 1488 cctgcctccgaggtgttaggggtgtggagaaaggggcagggcaggggtcagggatattg 1547
 Db 720 CCTGCCCTCCGAGGCTGTAGGGGTGTGGGAGAAAGGGGCGAGGCTCAGGGATATTG 779

QY 1548 agtgactccttggagctgtgggt 1607
 Db 780 AGTGACTGCTTGGAGTCTGGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 839

QY 1608 tcatcgctctgt 1667
 Db 840 TCATCGTCTCTGT 898

QY 1668 caggagggagagcagacgac 1687
 Db 899 CAGGAGGAGGAGCAGACGAC 918

RESULT 15
 HSNCF5
 LOCUS
 DEFINITION Homo sapiens p47-phox (NCF1) pseudogene, clone P39, exons 6 and 7.
 ACCESSION U72360
 VERSION U72360.1 GI:2754740
 KEYWORDS
 SEGMENT 5 of 6

SOURCE human.
ORGANISM Homo sapiens
REFERENCE 1 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS 1 (bases 1 to 904)
TITLE A p47-phox pseudogene carries the most common mutation causing p47-phox- deficient chronic granulomatous disease
J. Clin. Invest. 100 (8), 1907-1918 (1997)
JOURNAL 97474758
MEDLINE 2 (bases 1 to 904)
AUTHORS Chanock, S.J., Roessler, J., Hopkins, P.J., Lee, P.L., Bassett, D.T., Christensen, B., Curnutte, J.T. and Gorlach, A.
TITLE Characterization of the genomic structure of the p47-phox gene
JOURNAL Unpublished
AUTHORS 3 (bases 1 to 904)
AUTHORS Gorlach, A., Lee, P.L., Chanock, S.J. and Curnutte, J.T.
TITLE Direct Submission
JOURNAL Submitted (24-SEP-1996) Molecular and Experimental Medicine, The Scripps Research Institute, 10550 N. Torrey Pines Rd., CALI, La Jolla, CA 92037
FEATURES
source Location/Qualifiers
1..904
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="7"
/map="7q11.23"
/clone="P39"
/note="pseudogene is defined by the presence of a GT deletion at the beginning of exon 2"
1..123
/gene="NCF1"
/number=6
/pseudo
586..693
/gene="NCF1"
/number=7
/pseudo
BASE COUNT 164 a 251 c 293 g 196 t
ORIGIN
Query Match 25.9%; Score 870.8; DB 9; Length 904;
Best Local Similarity 99.0%; Pred. No. 8.1e-213;
Matches 897; Conservative 0; Mismatches 7; Indels 2; Gaps 2;
Qy 788 acataccggcccaatcctgcagacgtaccgcgcattgcgcgactacagagacct 847
Db 1 ACATCAGCGGCCCATCATCTCGAGAGCTACCGCGCCATTGCCAACTACGAGAAGACCT 60
Qy 848 cgggctccagatgctctgtccacggggagcgtagtgaggctcgtgagagagcgaga 907
Db 61 CGGGCTCCGAGATGGCTCTGTCCACGGGGGACGTGGTGGAGGTCGTGGAGAAGCGGAGA 120
Qy 908 gcggtcagacctccacaccttaacgggctcctccctggtgcctcaggaacccacagccac 967
Db 121 CGGGTCAGACCTCCACCTTACGGGGCTCTCTCCCTCGTCTCAGGAACCCACAGCCAC 180
Qy 968 aagccccctgcaaggctcaggagcctggccccctgggagactccagctctgttagggg 1027
Db 181 AAGCCCCCTGCCAA-GCTCAGGACAGCTGGCCCCCTGGGAGGACTCAGGCTCTGTTAGGGG 239
Qy 1028 ccctaaatgtctcccaactgtgggtgcgcttctctcttagtgtgacccctgtggtgg 1087
Db 240 CCCTAAATGTCTCCCCACATGTGGGTGGCTCTCTCTTAGTGTGACCCCTGTGGTGG 299
Qy 1088 ctgtgggcatctgtgcagggcgggcgggcgatgctgcgtgtctgtctgtggtg 1147
Db 300 CTGTGGGATCTGTGCATGGGAGCGCGGGGGGCATGTCTGCGGTCTGTCTGGATG 359
Qy 1148 ggtatgggacgctgtttcattatgaagtgggctcagagctgtgattctgtgagcatgtg 1207
|||||

Db 360 GGATGGGACCGCTCTGTTTCATTATGAAGTGGGCTCAGAGCTGTGATTCTGTGAGCATGTG 419
Qy 1208 tgcagcatgcatgacacctcatgtccagtgtagtgagtgacatttccaaatctgaag 1267
|||||
Db 420 TGCATGCATGATGACCTCATTTGCCAGTGTGTGAGGTGACATTTCCAAATCTGAG 479
Qy 1268 cattggacatcagtgctgtgtccctgtgtccctcaccatccctcctgaggtgcaggag 1327
|||||
Db 480 CATTGGACATCAGTGTGTGTGACCCCTGTGTCTCACCATCCCTGATGGCTGCAGGGAG 539
Qy 1328 ccgctgggccccctgcctcagtcacattcccgcacctctggcacagggttggtctctgt 1387
|||||
Db 540 CCGTGGGCCCTGCCCTCAGTCACATTTCCGACACTCTGGCACAGTTGGTGTCTGT 599
Qy 1388 cagatgaagcaaacgaggtgtggtccagcatccttccctcgagccctgagacagtcct 1447
|||||
Db 600 CAGATGAAGCAAAAGCGAGGCTGGATCCAGCATCTTCTCGAGCCCTCGACAGTCTCT 659
Qy 1448 gacgagacggaagaccctgagcccccaactatgcagdtgccccctgcctccagggtctag 1507
|||||
Db 660 GACGAGACGGAAGACCCCTGAGCCCAACTATGCAGGTGCCCTGCCAGGGCTGTAG 719
Qy 1508 ggggtgggagaaagggcagggcaggggtcaggggtatattgagtgactgtcttggaggtctg 1567
|||||
Db 720 GGGTGTGGGAGAAAGGGGCGAGGCTCAGGGCTATTTGAGTGACTGCTTTTGGAGTCTG 779
Qy 1568 ggcctgggtgctggcttgagagaaagtcagggctaaagatctcatcggtcttggtgggg 1627
|||||
Db 780 GGTGGTGTGCTGGCTTGGCAGAAAGTCAAGGCTAAGATCTCATCTCTCTGGCTTGGGG 839
Qy 1628 gccctggcaggttgtgtgcttgccttgggtctgagacagggaaccaggaggaggagcagacgac 1687
|||||
Db 840 GCCCTGGCAGGTTGTGTGATGCCCTTGGTCTGGACAGGAA-CAGGAGGAGGAGCAGCAGCAC 898
Qy 1688 tcggga 1693
|||||
Db 899 TCGGGA 904

Search completed: January 11, 2002, 17:18:54
Job time: 29486 sec

Parent case of 10/76734/

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 11, 2002, 11:35:07; Search time 8779.43 Seconds
(without alignments)
4113.771 Million cell updates/sec

Title: US-09-820-005-3_COPY_10200_13560
Perfect score: 3361
Sequence: 1 aaaggtcaggcactgtggc.....ctgtccaagtgggtgcatt 3361

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0
Searched: 11351937 seqs, 5372889281 residues
Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estfun:*
2: em_esthum:*
3: em_estin:*
4: em_estom:*
5: em_estpl:*
6: em_estba:*
7: em_estro:*
8: em_estov:*
9: em_hic:*
10: gb_est1:*
11: gb_est2:*
12: gb_hic:*
13: gb_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_inv:*
17: em_gss_pln:*
18: em_gss_pro:*
19: em_gss_rod:*
20: em_gss_vrt:*
21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	673.2	20.0	774	11	BG685989
2	362	10.8	15970	13	AQ839852
3	355.2	10.6	459	11	BF901739
4	350.2	10.4	743	10	AU130725
5	348.8	10.4	828	10	AL527073
6	348.4	10.4	732	13	AQ890095
7	347.4	10.3	1501	12	BC007465
8	345	10.3	658	13	AQ393450
9	342.2	10.2	680	13	AQ356404
10	340.4	10.1	372	11	BF900703
11	336.8	10.0	948	11	BG260565
12	335.6	10.0	877	13	AQ739838

13	332.2	9.9	947	11	BG680848
14	327.6	9.7	746	11	BF974962
15	327.4	9.7	710	13	AG010073
16	326	9.7	677	11	BG032943
17	324.6	9.7	778	10	AV764490
18	324.4	9.7	877	13	AQ739838
19	323.2	9.6	634	13	AQ543763
20	323.2	9.6	729	10	AV762129
21	322.6	9.6	705	10	AV760497
22	322.4	9.6	610	13	AQ479821
23	322.4	9.6	861	11	BG386803
24	321.8	9.6	679	11	BF346320
25	321.2	9.6	631	11	BE796439
26	320	9.5	1345	10	AV762220
27	318.6	9.5	728	10	AV763952
28	318.2	9.5	1345	10	AV762220
29	317.4	9.4	735	10	AL602867
30	316.8	9.4	851	10	AU122155
31	316.6	9.4	800	10	AU120942
32	316.4	9.4	922	10	AL515875
33	315.8	9.4	858	10	AU119532
34	315.6	9.4	948	11	BG260565
35	314.6	9.4	780	10	BE538259
36	314.4	9.4	6146	13	AQ839831
37	312.8	9.3	822	13	AQ748222
38	312.6	9.3	740	13	AQ878530
39	312.4	9.3	658	13	AQ393450
40	312.4	9.3	839	10	AV700988
41	312	9.3	1875	12	BC009270
42	311.8	9.3	759	10	AW976010
43	311.4	9.3	711	10	AV714931
44	310.4	9.2	891	11	BG744641
45	310	9.2	711	10	AL044340

ALIGNMENTS

RESULT 1
LOCUS BG685989 774 bp mRNA EST 01-MAY-2001
DEFINITION 602638552F1 NIH_MGC_48 Homo sapiens cDNA clone IMAGE:4766245 5', mRNA sequence.
ACCESSION BG685989
VERSION BG685989.1 GI:13917386
KEYWORDS EST.
SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 774)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaabs-remail.nih.gov

Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: L16M1626 row: g column: 14
High quality sequence stop: 773.
Location/Qualifiers
1. 774

FEATURES
source

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4766245"
/clone_lib="NIH_MGC_48"
/tissue_type="primary B-cells from tonsils (cell line)"
/lab_host="DH10B (phage-resistant)"

[illegible]

Db 7351 TGAGCCAAAGCGCCAGAC 7369

RESULT 6

US-08-814-095-7/c

; Sequence 7, Application US/08814095

; Patent No. 6025183

; GENERAL INFORMATION:

; APPLICANT: Soreq, Hermona

; APPLICANT: Zakut, Haim

; APPLICANT: Shani, Moshe

; TITLE OF INVENTION: TRANSGENIC ANIMAL ASSAY SYSTEM FOR

; TITLE OF INVENTION: ANTI-CHOLINESTERASE SUBSTANCES

; NUMBER OF SEQUENCES: 7

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: KOHN & ASSOCIATES

; STREET: 30500 No. 6025183Western Highway, Suite 410

; CITY: Farmington Hills

; STATE: Michigan

; COUNTRY: U.S.

; ZIP: 48334

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/814,095

; FILING DATE:

; CLASSIFICATION: 800

; ATTORNEY/AGENT INFORMATION:

; NAME: Montgomery, Ilene N.

; REGISTRATION NUMBER: 38,972

; REFERENCE/DOCKET NUMBER: 2391.00066

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (248) 539-5050

; TELEFAX: (248) 539-5055

; INFORMATION FOR SEQ ID NO: 7:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 35060 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

; MOLECULE TYPE: other nucleic acid

; DESCRIPTION: /desc = "Cosmid including ACHE

; DESCRIPTION: promotor, ACHE gene and ARS gene"

; HYPOTHETICAL: NO

; ANTI-SENSE: NO

; ORIGINAL SOURCE:

; ORGANISM: Homo sapiens

; POSITION IN GENOME:

; CHROMOSOME/SEGMENT: 7q22

; FEATURE:

; NAME/KEY: promotor

; LOCATION: 4089..22464

; OTHER INFORMATION: /function= "ACHE promotor"

; OTHER INFORMATION: /standard_name= "ACHE Promotor"

; FEATURE:

; NAME/KEY: exon

; LOCATION: 22465..22537

; OTHER INFORMATION: /function= "non-translated"

; OTHER INFORMATION: /gene= "ACHE"

; OTHER INFORMATION: /number= 1

; FEATURE:

; NAME/KEY: exon

; LOCATION: 24090..25177

; IDENTIFICATION METHOD: experimental

; OTHER INFORMATION: /function= "(translation start:

; OTHER INFORMATION: 24110)"

; OTHER INFORMATION: /evidence= EXPERIMENTAL

; OTHER INFORMATION: /gene= "ACHE"

; OTHER INFORMATION: /number= 2

; FEATURE:

; NAME/KEY: exon

; LOCATION: 25524..26009

; IDENTIFICATION METHOD: experimental

; OTHER INFORMATION: /evidence= EXPERIMENTAL

; OTHER INFORMATION: /gene= "ACHE"

; OTHER INFORMATION: /number= 3

; FEATURE:

; NAME/KEY: exon

; LOCATION: 27005..27274

; IDENTIFICATION METHOD: experimental

; OTHER INFORMATION: /evidence= EXPERIMENTAL

; OTHER INFORMATION: /gene= "ACHE"

; OTHER INFORMATION: /number= 4

; FEATURE:

; NAME/KEY: exon

; LOCATION: 27255..28007

; IDENTIFICATION METHOD: experimental

; OTHER INFORMATION: /evidence= EXPERIMENTAL

; OTHER INFORMATION: /gene= "ACHE"

; OTHER INFORMATION: /number= 5

; FEATURE:

; NAME/KEY: terminator

; LOCATION: 27385..27387

; FEATURE:

; NAME/KEY: exon

; LOCATION: 28008..28129

; IDENTIFICATION METHOD: experimental

; OTHER INFORMATION: /evidence= EXPERIMENTAL

; OTHER INFORMATION: /gene= "ACHE"

; OTHER INFORMATION: /number= 6

; FEATURE:

; NAME/KEY: terminator

; LOCATION: 28129..28131

; FEATURE:

; NAME/KEY: exon

; LOCATION: complement (34528..34895)

; OTHER INFORMATION: /function= "arsenite resistance

; OTHER INFORMATION: gene"

; OTHER INFORMATION: /gene= "AR"

; OTHER INFORMATION: /number= 1

; FEATURE:

; NAME/KEY: exon

; LOCATION: complement (34092..34358)

; OTHER INFORMATION: /gene= "AR"

; OTHER INFORMATION: /number= 2

; FEATURE:

; NAME/KEY: exon

; LOCATION: complement (33779..33963)

; OTHER INFORMATION: /gene= "AR"

; OTHER INFORMATION: /number= 3

; FEATURE:

; NAME/KEY: exon

; LOCATION: complement (33493..33591)

; OTHER INFORMATION: /gene= "AR"

; OTHER INFORMATION: /number= 4

; FEATURE:

; NAME/KEY: exon

; LOCATION: complement (33297..33408)

; OTHER INFORMATION: /gene= "AR"

; OTHER INFORMATION: /number= 5

; FEATURE:

; NAME/KEY: exon

; LOCATION: complement (32959..33094)

; OTHER INFORMATION: /gene= "AR"

; OTHER INFORMATION: /number= 6

; FEATURE:

; NAME/KEY: exon

; LOCATION: complement (32569..32628)

; OTHER INFORMATION: /gene= "AR"

; OTHER INFORMATION: /number= 7

; FEATURE:

; NAME/KEY: exon

; LOCATION: complement (32386..32468)


```

;
; GENERAL INFORMATION:
; APPLICANT: Chader, Gerald J.; Becerra, S.
; APPLICANT: Patricia; Schwartz, Joan P.;
; APPLICANT: Taniwaki, Takayuki
; TITLE OF INVENTION: PIGMENT EPITHELIUM
; TITLE OF INVENTION: DERIVED FACTOR: CHARACTERIZATION OF ITS NOVEL
; TITLE OF INVENTION: BIOLOGICAL ACTIVITY AND SEQUENCES ENCODING
; TITLE OF INVENTION: AND EXPRESSING THE PROTEIN
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Morgan & Finnegan
; STREET: 345 Park Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy Disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/257,963B
; FILING DATE:
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/952,796
; FILING DATE: 24-SEPT-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: DOROTHY R. AUTH
; REGISTRATION NUMBER: 36434
; REFERENCE/DOCKET NUMBER: 20264126US1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 758-4800
; TELEFAX: (212) 751-6849
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4421 Base Pairs
; TYPE: Nucleic Acid
; STRANDEDNESS: Double
; TOPOLOGY: Unknown
; MOLECULE TYPE: Genomic DNA
; ORIGINAL SOURCE:
; ORGANISM: Human
; FEATURE:
; NAME/KEY: JT101
; LOCATION:
; IDENTIFICATION METHOD:
; OTHER INFORMATION: 7.1 kb Bam HI
; OTHER INFORMATION: fragment Derived from human placental
; OTHER INFORMATION: genomic DNA
;
; US-08 :57-963B-9
;
; Query/ Match 9.7%; Score 326.6; DB 2; Length 4421;
; Best Local Similarity 75.4%; Pred. No. 8.9e-69;
; Matches 500; Conservative 0; Mismatches 99; Indels 64; Gaps 5;
;
; Qy 1012 tatttatttacttttttgagatgagatctcactcccttgctgaaggtgagtgagtggtg 1971
; Db 3491 TATTTTCTTTTCTTTTGGAGATGAGTCTGGCTCTGTACCCAGGGCTGGAGTGCAGTGG 3550
;
; Qy 1072 tgcctatccagctcactgcaacctctgctccaggatcgaagtgtatctccagctcagc 2031
; Db 3451 CGCGATCTCGGCTCACTGCAAGCTCTTCTCCCGGGTTCACCCCATTTCTCTGCTCAAC 3610
;
; Qy 2032 ctccccaggtagctgggattacaagcatgcaccaccatgcctggataa--ttttttgatt 2089
; Db 3011 CTCCCGA-GTAGCTGTGAATACAGCGGTCGCCACCACCGCCGACTAATTTTTTTGTATT 3669
;
; Qy 2090 tttagtgagacgggggtttaccagggttgccagactgggtcgaacttttgacctcagg 2149
; Db 3070 TTTAGTAGAGACGGGGTTTACCGTGTAGCCAGGATGGTCTGGATCTCTTGACCTC--G 3727

```

```

;
; Qy 2150 tgatcgtcctcctcgtcctcccaagtgctggaattacagatgtgagccactgtccctg 2209
; Db 3728 TGATCTGCTGCTCGGCTCCCAAGTGTGGGATACAGGCTTGAGCCACCGCACCCG 3787
;
; Qy 2210 gctcgtggttaccacattttaaaatggagtgatttcaaccttttatgtggattacagctt 2269
; Db 3788 GCC----- 3790
;
; Qy 2270 gtttttttttttttttttttttttttttttttttttttttttttttttttttttttt 2329
; Db 3791 -TCATTATTTT----- 3849
;
; Qy 2330 gcaatctcagctcactgcaacctttagcctcctggttcaagcaattctcctgctcagcc 2389
; Db 3850 CGATCTCGGCTCACTGCAACCTCCGCTCTCTGGTTCAAGAGATTCTCCTGCTCAGCC 3909
;
; Qy 2390 acctgagtagcctggggttacagcatgcaccacagccagggctaattttttttttttt 2449
; Db 3910 TCCCAAGTAG-CTGGGATTACAGGTGCCACACACACCGCTGGCTAGTTTGTATTTT 3968
;
; Qy 2450 tagtagagatggggttttcgcccatttggccaggtgtgtctcgaactcctgacctcaggtg 2509
; Db 3969 TAGTAAAGATGGGTTTACCATGTTGCCAGGCTGGTCTTCAACTCCTGACATCAGGTG 4028
;
; Qy 2510 atccgcccgccttgccctcccaagtgctagattacaggtgggaacacaccttgcaccagc 2569
; Db 4029 ATCCGCCACCTTAGCTCCCAAAAGTCTGGGATTACAGGCGTGAGCCACCATACCTGGC 4088
;
; Qy 2570 ctg 2572
; Db 4089 CAG 4091
;
; RESULT 10
; PCT-US95-07201-9
; Sequence 9, Application PC/TUS9507201
; GENERAL INFORMATION:
; APPLICANT: Chader, Gerald J.; Becerra, Sofia
; APPLICANT: Patricia; Schwartz, Joan P.;
; APPLICANT: Taniwaki, Takayuki
; TITLE OF INVENTION: PIGMENT EPITHELIUM
; TITLE OF INVENTION: DERIVED FACTOR: CHARACTERIZATION GENOMIC
; TITLE OF INVENTION: ORGANIZATION AND SEQUENCE OF THE PEDF GENE
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Morgan & Finnegan, L.L.P.
; STREET: 345 Park Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy Disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/07201
; FILING DATE: 06-JUN-1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/367,841
; FILING DATE: 30-DEC-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/257,963
; FILING DATE: 07-JUN-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/952,796
; FILING DATE: 24-SEP-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: DOROTHY R. AUTH
; REGISTRATION NUMBER: 36434

```

```

Db 4089 CAG 4091

RESULT 11
US-09-009-913-1
; Sequence 1, Application US/09009913
; Patent No. 6087485
; GENERAL INFORMATION:
; APPLICANT: Axy's Pharmaceuticals, Inc.
; TITLE OF INVENTION: Asthma Related Genes
; NUMBER OF SEQUENCES: 339
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Bozicevic & Reed, LLP
; STREET: 285 Hamilton Ave, Suite 200
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94301
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA: US/09/009,913
; FILING DATE: 21-JAN-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Sherwood, Pamela J
; REGISTRATION NUMBER: 36,677
; REFERENCE/DOCKET NUMBER: SEQ-4P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-327-3231
; TELEFAX: 650-327-3231
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 72928 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
US-09-009-913-1

Query Match 9.6%; Score 321.8; DB 3; Length 72928;
Best Local Similarity 73.9%; Pred. No. 3.9e-67;
Matches 500; Conservative 0; Mismatches 122; Indels 55; Gaps 5

Qy 1924 ttttttgagatagagtctcaactcccttctaaggctgagtgcggtggtgtctatctcagc 1983
||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 47733 TTTcATGAGACAAGGCTCTGTCTGTTGCCAGGCTGGAGTCAGCTGGCGTGATCTCTGC 47792
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1984 tcactgcaacctctgcctcccaggatcaagtgaattctccagctcagcctcccccaggtag 2043
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 47793 TCACCTTCAACCTCTGCCCTCCAGGTTCAAGTCAGCTGCTCATGCTCAGGCTCTGGA-GTAG 47851
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 2044 ctgggattacaagcatgcaccaccatgcctggataattttttagtttgagaagg 2103
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 47852 CTGGGACTACAGGTGCGTGCCACCAACCTTGGCTAAATTTTGTATTTTGGTAGAGATGG 47911
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 2104 ggtttcaccaaggttggccagactggctcgaaactctgaactctgaactcaggtgatctgcctgctc 2163
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 47912 GGTATTGGCATGTTTGGCCAGGCTGGTCTCGAACTCCTGGCCCTCAAGTATTTGACTGGCT 47971
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 2164 cggcctcccaagtgctggaattacagatgtgagccactgtccctcgccctggttaccac 2223
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 47972 TGGGCTTCCCAAGTGCTGGAAATTACAGGCGGTGAGGCCACCATGCTCTGGGACTTGCTGA- 48026
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 2224 attttaaataagatgatttcaccccttttatgtggaatttacagactgtgttttttttttt 2283
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```


Search completed: January 11, 2002, 12:08:49
Job time: 10881 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 11, 2002, 09:07:38 ; Search time 25.69 seconds
(without alignments)
2197.786 Million cell updates/sec

Title: US-09-820-005-2
Perfect score: 2035
Sequence: 1 MGDFTIRHIALLGFERFVP.....ADILNRCSESTRKRLASAV 386

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5
Searched: 473505 seqs, 146272329 residues
Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL17:*
1: sp-archaea:*
2: sp-bacteria:*
3: sp-fungi:*
4: sp-human:*
5: sp-invertebrate:*
6: sp-mammal:*
7: sp-mhc:*
8: sp-organelle:*
9: sp-phage:*
10: sp-plant:*
11: sp-rodent:*
12: sp-virus:*
13: sp-vertebrate:*
14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2023	99.4	390	4 Q9BU90	Q9BU90 homo sapien
2	2015	99.0	390	4 O43842	O43842 homo sapien
3	2005.5	98.6	389	4 Q9UDV9	Q9UDV9 homo sapien
4	1717.5	84.4	391	6 Q9N088	Q9N088 turslops tr
5	1697	83.4	390	11 Q9J134	Q9J134 mus musculu
6	1689.5	83.0	389	11 Q9NM65	Q9NM65 rattus norv
7	1689	83.0	390	11 O70144	O70144 mus musculu
8	1411	69.3	276	4 Q9BX18	Q9BX18 homo sapien
9	1346.5	66.2	310	11 Q9JK56	Q9JK56 rattus norv
10	1290	63.4	254	4 Q9BX17	Q9BX17 homo sapien
11	504.5	24.8	1124	11 O89032	O89032 mus musculu
12	334.5	16.4	1031	4 Q9H462	Q9H462 homo sapien
13	275	13.5	940	4 O43302	O43302 homo sapien
14	197	9.7	339	6 Q9GM25	Q9GM25 turslops tr
15	188	9.2	1248	4 Q9NZM2	Q9NZM2 homo sapien
16	188	9.2	1676	4 Q9ULG4	Q9ULG4 homo sapien
17	188	9.2	1681	4 Q9NYG0	Q9NYG0 homo sapien
18	188	9.2	1696	4 Q9NZM3	Q9NZM3 homo sapien
19	185.5	9.1	550	4 Q9P2Q1	Q9P2Q1 homo sapien

20	183	9.0	464	4 O95062	O95062 homo sapien
21	179.5	8.8	1270	13 O42287	O42287 xenopus lae
22	178	8.7	348	4 Q9BU98	Q9BU98 homo sapien
23	177.5	8.7	1197	11 Q9Z0R5	Q9Z0R5 mus musculu
24	177.5	8.7	1658	11 Q9Z0R6	Q9Z0R6 mus musculu
25	176.5	8.7	248	4 Q15812	Q15812 homo sapien
26	172	8.5	1714	11 Q9Z0R4	Q9Z0R4 mus musculu
27	169.5	8.3	102	4 Q9NTM6	Q9NTM6 homo sapien
28	168.5	8.3	248	11 Q9R143	Q9R143 mus musculu
29	168	8.3	637	11 Q9JLQ0	Q9JLQ0 mus musculu
30	166.5	8.2	1217	11 Q9WVE9	Q9WVE9 rattus norv
31	166	8.2	1220	4 Q9UNK1	Q9UNK1 homo sapien
32	166	8.2	1220	4 Q9UET5	Q9UET5 homo sapien
33	166	8.2	1721	4 O95216	O95216 homo sapien
34	166	8.2	1721	4 O9UNK2	O9UNK2 homo sapien
35	164.5	8.1	639	4 Q9Y5K6	Q9Y5K6 homo sapien
36	164	8.1	239	11 Q9D747	Q9D747 mus musculu
37	164	8.1	1011	5 O61639	O61639 drosophila
38	161	7.9	641	11 O88903	O88903 mus musculu
39	161	7.9	1094	5 O61618	O61618 drosophila
40	160.5	7.9	1146	11 Q9WVE1	Q9WVE1 rattus norv
41	159.5	7.8	1097	5 Q9VIF7	Q9VIF7 drosophila
42	151	7.4	187	4 Q9UK60	Q9UK60 homo sapien
43	150.5	7.4	665	4 Q9NYR0	Q9NYR0 homo sapien
44	150	7.4	585	11 Q9JKQ1	Q9JKQ1 rattus norv
45	147.5	7.2	290	3 P78815	P78815 schizosacch

ALIGNMENTS

RESULT 1
Q9BU90
ID Q9BU90 PRELIMINARY: PRT: 390 AA.
AC Q9BU90;
DT 01-JUN-2001 (TremBLrel. 17, Created)
DT 01-JUN-2001 (TremBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TremBLrel. 17, Last annotation update)
DE NEUTROPHIL CYTOSOLIC FACTOR 1 (47KD, CHRONIC GRANULOMATOUS DISEASE, DEUTOSOMAL 1).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LYMPHOMA;
RA Strausberg R.;
RL Submitted (FEB-2001) to the EMBL/GenBank/DDBJ databases.
DR EMBL: BC002815; AAH02816.1; ...
SQ SEQUENCE 390 AA; 44681 MW; 3D91EDC99A1B6417 CRC64;

Query Match 99.4%; Score 2023; DB 4; Length 390;
Best Local Similarity 99.0%; Pred No. 4.2e-153;
Matches 386; Conservative 0; Mismatches 0; Indels 4; Gaps 1;

Qy	1	MGDTFIRHIALLGFERFVP	Q9BU90	Q9BU90	60
Db	1	MGDTFIRHIALLGFERFVP	Q9BU90	Q9BU90	60
Qy	61	EGAGINPENRIIPHLPAKPFQDQRAAENRQGLTEYCS	Q9BU90	Q9BU90	120
Db	61	EGAGINPENRIIPHLPAKPFQDQRAAENRQGLTEYCS	Q9BU90	Q9BU90	120
Qy	121	RPDDLKLPDNGTKPEYLMKPKCKSTATDITGPILOTYRAIANYEKTSSEWALSTG	Q9BU90	Q9BU90	180
Db	121	RPDDLKLPDNGTKPEYLMKPKCKSTATDITGPILOTYRAIANYEKTSSEWALSTG	Q9BU90	Q9BU90	180
Qy	181	DVVEVVEKSESGWFCQMKAKRGWIPASPLEPDSDEDEPVPYVAIKAYTAV	Q9BU90	Q9BU90	240
Db	181	DVVEVVEKSESGWFCQMKAKRGWIPASPLEPDSDEDEPVPYVAIKAYTAV	Q9BU90	Q9BU90	240

```
QY 241 EGDEVSLLGEAVEVIHKLDDGW----KDDVTGYPPSMYLOKSGDVSAQAQRIKRGAPP 296
DB 241 EGDEVSLLGEAVEVIHKLDDGWVIRKDDVTGYPPSMYLOKSGDVSAQAQRIKRGAPP 300
QY 297 RRSSTRNHSHIORSKRKLSQDAYRRNSVRFLOQRRRQARPQSPGSPLEEROTQRSK 356
DB 301 RRSSTRNHSHIORSKRKLSQDAYRRNSVRFLOQRRRQARPQSPGSPLEEROTQRSK 360
QY 357 POPAVPPRPSADLIILNRCSESTKRKLASAV 386
DB 361 POPAVPPRPSADLIILNRCSESTKRKLASAV 390

RESULT 2
O43842
ID O43842 PRELIMINARY; PRT; 390 AA.
AC O43842;
DT 01-JUN-1998 (TremBLrel. 06, Created)
DT 01-JUN-1998 (TremBLrel. 06, Last sequence update)
DT 01-JUN-2001 (TremBLrel. 17, Last annotation update)
DE P47-PHOX.
GN NCF1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97474758; PubMed=9329953;
RA Gorlach A., Lee P.L., Roesler J., Hopkins P.J., Christensen B.,
RA Green E.D., Chanock S.J., Curnutte J.T.;
RT "A p47-phox pseudogene carries the most common mutation causing p47-
RT phox-deficient chronic granulomatous disease.";
RL J. Clin. Invest. 100:1907-1918(1997).
RN [2]
RP SEQUENCE FROM N.A.
RA Chanock S.J., Roesler J., Zhan S., Hopkins P., Lee P., Barrett D.,
RA Christensen B.L., Curnutte J.T., Goerlach A.;
RT "Genomic structure of the human p47-phox gene.";
RL Blood Cells Mol. Dis. 0:0-0(2000).
DR EMBL; U57835; AAB95193.1; -.
DR EMBL; U57833; AAB95193.1; JOINED.
DR EMBL; U57834; AAB95193.1; JOINED.
DR EMBL; AF184614; AAF34737.1; -.
DR InterPro; IPR001655; P47PHOX.
DR InterPro; IPR001683; PX.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00787; PX; 1.
DR Pfam; PF00018; SH3; 2.
DR PRINTS; PR00498; P47PHOX.
DR SMART; SM00312; PX; 1.
DR SMART; SM00326; SH3; 2.
DR PROSITE; PS00002; SH3; 2.
SQ SEQUENCE 390 AA; 44695 MW; A19746E845B044C8 CRC64;

Query Match 99.0%; Score 2015; DB 4; Length 390;
Best Local Similarity 98.5%; Pred. No. 1.8e-152;
Matches 384; Conservative 1; Mismatches 1; Indels 4; Gaps 1;

QY 1 MGDTRIRHALLGFKEKRFVPSQHYVMFLVKWQDLSEKVVYRFRTEIYEFHKLKEMPFI 60
DB 1 MGDTRIRHALLGFKEKRFVPSQHYVMFLVKWQDLSEKVVYRFRTEIYEFHKLKEMPFI 60
QY 61 EAGAINPENRIIPHPAPKWFQDQRAAENRQGLTEYCSLMSLPTKISRCPHLLDFFKV 120
DB 61 EAGAINPENRIIPHPAPKWFQDQRAAENRQGLTEYCSLMSLPTKISRCPHLLDFFKV 120
QY 121 RPDDLKLPDNTQTKPKETLMPKDGKSTATDITGPIILQTYRAIANYEKTSSEMASTG 180
DB 121 RPDDLKLPDNTQTKPKETLMPKDGKSTATDITGPIILQTYRAIANYEKTSSEMASTG 180
QY 181 DVVEVSESGWFCQMKAKRGWIPASFLPDSDETEDEPNYAGEPYVAIKAYTAV 240
```

```
DB 181 DVVEVSESGWFCQMKAKRGWIPASFLPDSDETEDEPNYAGEPYVAIKAYTAV 240
QY 241 EGDEVSLLGEAVEVIHKLDDGW----KDDVTGYPPSMYLOKSGDVSAQAQRIKRGAPP 296
DB 241 EGDEVSLLGEAVEVIHKLDDGWVIRKDDVTGYPPSMYLOKSGDVSAQAQRIKRGAPP 300
QY 297 RRSSTRNHSHIORSKRKLSQDAYRRNSVRFLOQRRRQARPQSPGSPLEEROTQRSK 356
DB 301 RRSSTRNHSHIORSKRKLSQDAYRRNSVRFLOQRRRQARPQSPGSPLEEROTQRSK 360
QY 357 POPAVPPRPSADLIILNRCSESTKRKLASAV 386
DB 361 POPAVPPRPSADLIILNRCSESTKRKLASAV 390

RESULT 3
O9UDV9
ID O9UDV9 PRELIMINARY; PRT; 389 AA.
AC O9UDV9;
DT 01-MAY-2000 (TremBLrel. 13, Created)
DT 01-MAY-2000 (TremBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TremBLrel. 17, Last annotation update)
DE WUGSC:H_DJ0771P04.4 PROTEIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99063792; PubMed=9847074;
RA Sulston J.E., Waterston R.;
RT "Toward a complete human genome sequence.";
RL Genome Res. 8:1097-1108(1998).
RN [2]
RP SEQUENCE FROM N.A.
RA Kalicki J., Lapiant Y.;
RT "The sequence of Homo sapiens PAC clone RP4-771P4.";
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Waterston R.;
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Waterston R.;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC004883; AAD15422.1; -.
DR InterPro; IPR001452; SH3.
DR InterPro; IPR001655; P47PHOX.
DR InterPro; IPR001683; PX.
DR Pfam; PF00018; SH3; 2.
DR Pfam; PF00787; PX; 1.
DR PRINTS; PR00498; P47PHOX.
DR PRINTS; PR00452; SH3DOMAIN.
DR PROSITE; PS00002; SH3; 2.
DR SMART; SM00312; PX; 1.
DR SMART; SM00326; SH3; 2.
SQ SEQUENCE 389 AA; 44584 MW; 51BF3DA576497EEC CRC64;

Query Match 98.6%; Score 2005.5; DB 4; Length 389;
Best Local Similarity 98.7%; Pred. No. 1e-151;
Matches 385; Conservative 0; Mismatches 0; Indels 5; Gaps 2;

QY 1 MGDTRIRHALLGFKEKRFVPSQHYVMFLVKWQDLSEKVVYRFRTEIYEFHKLKEMPFI 60
DB 1 MGDTRIRHALLGFKEKRFVPSQHYVMFLVKWQDLSEKVVYRFRTEIYEFHKLKEMPFI 60
QY 61 EAGAINPENRIIPHPAPKWFQDQRAAENRQGLTEYCSLMSLPTKISRCPHLLDFFKV 120
DB 61 EAGAINPENRIIPHPAPKWFQDQRAAENRQGLTEYCSLMSLPTKISRCPHLLDFFKV 120
```

```

QY 121 RDDLKLPDNTQKPEYLMFKDCKSTATDITGPIILQTYRAIANYEKTSSEMALSTG 180
DB 121 RDDLKLPDNTQKPEYLMFKDCKSTATDITGPIILQTYRAIANYEKTSSEMALSTG 180
QY 181 DVVEVVEKSESGWFCOMKAKRGWIPASFLPLDSDPDETPENYAGEPYVAIKAYTAV 240
DB 181 DVVEVVEKSESGWFCOMKAKRGWIPASFLPLDSDPDETPENYAGEPYVAIKAYTAV 240
QY 241 EGDEVSLLGEAVEVIHKLDDGW----KDDVTGYFPSPMYLQKSGDVSOAQRIKRGAPP 296
DB 241 EGDEVSLLGEAVEVIHKLDDGW----KDDVTGYFPSPMYLQKSGDVSOAQRIKRGAPP 296
QY 297 RRSIRNAHSIHQRKRKLSQDAYRRNSVRFLOQRRRQARPGPQSPGSPLEEROTQSK 356
DB 297 RRSIRNAHSIHQRKRKLSQDAYRRNSVRFLOQRRRQARPGPQSPGSPLEEROTQSK 356
QY 357 POPAVPPRPSADLILNRCSESTRKRLASAV 386
DB 360 POPAVPPRPSADLILNRCSESTRKRLASAV 389

RESULT 4
Q9N0E8 PRELIMINARY: PRT: 391 AA.
AC Q9N0E8
DT 01-OCT-2000 (TremBLrel. 15, Created)
DT 01-OCT-2000 (TremBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TremBLrel. 17, Last annotation update)
DE P47-PHOX.
GN Tursiops truncatus (Atlantic bottle-nosed dolphin).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Cetacea; Odontoceti; Delphinidae;
OC Tursiops.
OX NCBI_TaxID=9739;
RN [1]
RP SEQUENCE FROM N.A.
RT Inoue Y., Ito T., Sakai T.;
RT "Molecular Cloning and Identification of Bottle-Nosed Dolphin
RT P47phox.";
RL Submitted (D8C-1999) to the EMBL/GenBank/DBJ databases.
RL EMBL; AB035594; BAA36544.1;
DR InterPro; IPR001655; P47PHOX.
DR InterPro; IPR001683; PX.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00787; PX; 1.
DR Pfam; PF00018; SH3; 2.
DR PRINTS; PR00498; P47PHOX.
DR PRINTS; PR00452; SH3DOMAIN.
DR SMART; SM00312; PX; 1.
DR SMART; SM00326; SH3; 2.
DR PROSITE; PS50002; SH3; 2.
SQ SEQUENCE 391 AA; 44766 MW; F09E96D6B74B7C7C CRC64;

Query Match 84.4%; Score 1717.5; DB 6; Length 391;
Best Local Similarity 82.7%; Pred. No. 8.8e-129;
Matches 324; Conservative 30; Mismatches 31; Indels 7; Gaps 3;

QY 1 MGDFTIRHIALGFEKRVPSQHYVMFLVKWQDLSEKVVYRRFTIYEFHFKLKEPFI 60
DB 1 MGDFTIRHIALGFEKRVPSQHYVMFLVKWQDLSEKVVYRRFTIYEFHFKLKEPFI 60

QY 61 EAGATNPENRIIPHPAPKWFQDQRAAENRQGLTEYCYSTLMSLPTKISRCCHLLDFKVV 120
DB 61 EAGATNPENRIIPHPAPKWFQDQRAAENRQGLTEYCYSTLMSLPTKISRCCHLLDFKVV 120

QY 121 RDDLKLPDNTQKPEYLMFKDCKSTATDITGPIILQTYRAIANYEKTSSEMALSTG 180
DB 121 RDDLKLPDNTQKPEYLMFKDCKSTATDITGPIILQTYRAIANYEKTSSEMALSTG 180

QY 181 DVVEVVEKSESGWFCOMKAKRGWIPASFLPLDSDPDETPENYAGEPYVAIKAYTAV 240
DB 181 DVVEVVEKSESGWFCOMKAKRGWIPASFLPLDSDPDETPENYAGEPYVAIKAYTAV 240

```

```

DB 181 DVVDVVEKSESGWFCOMKTKRGWVPASYLEPLDSPDEADPEPNYEGEPTIIEAYTAV 240
QY 241 EGDEVSLLGEAVEVIHKLDDGW----KDDVTGYFPSPMYLQKSGDVSOAQRIKRGAPP 296
DB 241 EGDEVSLLGEAVEVIHKLDDGW----KDDVTGYFPSPMYLQKSGDVSOAQRIKRGAPP 296
QY 297 RRSIRNAHSIHQRKRKLSQDAYRRNSVRFLOQRRRQARPGPQSPGSPLEER--QTOR 354
DB 297 RRSIRNAHSIHQRKRKLSQDAYRRNSVRFLOQRRRQARPGPQSPGSPLEER--QTOR 354
QY 355 SKPQAVPPRPSADLILNRCSESTRKRLASAV 386
DB 360 PKPQAVPPRPSADLILNRCSESTRKRLASAV 391

RESULT 5
Q9JI34 PRELIMINARY: PRT: 390 AA.
AC Q9JI34
DT 01-OCT-2000 (TremBLrel. 15, Created)
DT 01-OCT-2000 (TremBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TremBLrel. 17, Last annotation update)
DE P47-PHOX.
GN Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RT Green E.D.;
RT Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
RL EMBL; AF267747; AAF30134.1;
DR MGD; MGI:97283; Ncfl.
DR InterPro; IPR001655; P47PHOX.
DR InterPro; IPR001683; PX.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00787; PX; 1.
DR Pfam; PF00018; SH3; 2.
DR PRINTS; PR00498; P47PHOX.
DR PRINTS; PR00452; SH3DOMAIN.
DR SMART; SM00312; PX; 1.
DR SMART; SM00326; SH3; 2.
DR PROSITE; PS50002; SH3; 2.
SQ SEQUENCE 390 AA; 44667 MW; C8EFAB953839CE9A CRC64;

Query Match 83.4%; Score 1697; DB 11; Length 390;
Best Local Similarity 81.1%; Pred. No. 3.8e-127;
Matches 317; Conservative 34; Mismatches 34; Indels 6; Gaps 3;

QY 1 MGDFTIRHIALGFEKRVPSQHYVMFLVKWQDLSEKVVYRRFTIYEFHFKLKEPFI 60
DB 1 MGDFTIRHIALGFEKRVPSQHYVMFLVKWQDLSEKVVYRRFTIYEFHFKLKEPFI 60

QY 61 EAGATNPENRIIPHPAPKWFQDQRAAENRQGLTEYCYSTLMSLPTKISRCCHLLDFKVV 120
DB 61 EAGATNPENRIIPHPAPKWFQDQRAAENRQGLTEYCYSTLMSLPTKISRCCHLLDFKVV 120

QY 121 RDDLKLPDNTQKPEYLMFKDCKSTATDITGPIILQTYRAIANYEKTSSEMALSTG 180
DB 121 RDDLKLPDNTQKPEYLMFKDCKSTATDITGPIILQTYRAIANYEKTSSEMALSTG 180

QY 181 DVVEVVEKSESGWFCOMKAKRGWIPASFLPLDSDPDETPENYAGEPYVAIKAYTAV 240
DB 181 DVVEVVEKSESGWFCOMKAKRGWIPASFLPLDSDPDETPENYAGEPYVAIKAYTAV 240

```

Db 301 PRSTRIRNAQSIHQSRKRLSODTYRRNSVRFLOQRRRGRGPLSTDG-TKDNFSTPRV 359
QY 356 KQPAVPPRPSADLIINRCSESTKRLASAV 386
DB 360 KQPAVPPRPSDDLILHRCSTKRLKLTSAV 390
RESULT 6
Q99M65 PRELIMINARY; PRT; 389 AA.
ID Q99M65;
AC Q99M65;
DT 01-JUN-2001 (TReMBLrel. 17, Created)
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE P47 PHOX.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY;
RA Tanabe M., Radmark O.P.;
RT "Full length sequence of rat p47 phox.";
RL Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AY029167; AAK31797.1; -.
SQ SEQUENCE 389 AA; 44742 MW; BFC75842E53E68A4 CRC64;

Query Match 83.0%; Score 1689.5; DB 11; Length 389;
Best Local Similarity 81.1%; Pred. No. 1.5e-126;
Matches 317; Conservative 34; Mismatches 33; Indels 7; Gaps 3;
QY 1 MGDFTIRHIALGFEKRFVPSOHYVYMFVLVKWQDLSEKVVYRRFTIYEFHKLKEMFPI 60
DB 1 MGDFTIRHIALGFEKRFVPSOHYVYMFVLVKWQDLSEKVVYRRFTIYEFHKLKEMFPI 60
QY 61 EAGAINPENRIIPLPAKFWDCQRAAENRQGLTEYCTLSLPTKISRCPHLLDFFKV 120
DB 61 EAGEIHTENRVIPHLPAKFWDCQRAAENRQGLTEYCTLSLPTKISRCPHLLDFFKV 120
QY 121 RPDDLKLPDNTQKPEYLMKDGKSTATDITGPILQTYRAIANYEKTSSEWALSTG 180
DB 121 RPDDLKLPDNTQKPEYLMKDGKSTATDITGPILQTYRAIANYEKTSSEWALSTG 180
QY 181 DVVEVVEKSESGWFCOMKAKRGWIPASPLEDSDPEDEPNNYAGPYVAIKAYTAV 240
DB 181 DVVDVVEKSESGWFCOMKAKRGWIPASPLEDSDPEDEPNNYAGPYVAIKAYTAV 240
QY 241 EGDEVSLLEGEAVEVTHKLLDGW---KDDVTGYFPSPMYLQKSGQDVSOAQRQIK-RGAP 295
DB 241 EEDVSLSEGEAVEVTHKLLDGW---KDDVTGYFPSPMYLQKSGQDVSOAQRQIK-RGAP 295
QY 296 PRSSIRNAHSTHQSRKRLSODAYRRNSVRFLOQRRRQARPGPSGSPLEERQOTRS 355
DB 301 PRSTRIRNAQSIHQSRKRLSODTYRRNSVRFLOQRRRGRGPLSTDG--KDNFSTPRA 358
QY 356 KQPAVPPRPSADLIINRCSESTKRLASAV 386
DB 359 KQPAVPPRPSDDLILHRCSTKRLKLTSAV 389

RESULT 7
Q99M65 PRELIMINARY; PRT; 390 AA.
ID Q99M65;
AC Q99M65;
DT 01-AUG-1998 (TReMBLrel. 07, Created)
DT 01-AUG-1998 (TReMBLrel. 07, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE P47 PHOX.
GN NCFL.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LEUKEMIA;
RX MEDLINE=98149672; PubMed=9490028;
RA Mizuki K., Kadomatsu K., Hata K., Ito T., Fan Q.-W., Kage Y.,
RA Fukumaki Y., Sakaki Y., Takeshige K., Sumimoto H.;
RT "Functional modules and expression of mouse p40(phox) and p67(phox),
RT SH3-domain-containing proteins involved in the phagocyte NADPH oxidase
RT complex.";
RL Eur. J. Biochem. 251:573-582(1998).
DR EMBL; AB026663; BAA35649.1; -.
DR MGD; MGI:97283; Ncfl.
DR InterPro; IPR001655; P47PHOX.
DR InterPro; IPR001883; PX.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00787; PX; 1.
DR Pfam; PF00018; SH3; 2.
DR PRINTS; PR00498; P47PHOX.
DR PRINTS; PR00452; SH3DOMAIN.
DR SMART; SM00312; PX; 1.
DR SMART; SM00326; SH3; 2.
DR PROSITE; PS00002; SH3; 2.
SQ SEQUENCE 390 AA; 44698 MW; A383DB953839CFBC CRC64;
Query Match 83.0%; Score 1689; DB 11; Length 390;
Best Local Similarity 80.8%; Pred. No. 1.6e-126;
Matches 316; Conservative 34; Mismatches 35; Indels 6; Gaps 3;
QY 1 MGDFTIRHIALGFEKRFVPSOHYVYMFVLVKWQDLSEKVVYRRFTIYEFHKLKEMFPI 60
DB 1 MGDFTIRHIALGFEKRFVPSOHYVYMFVLVKWQDLSEKVVYRRFTIYEFHKLKEMFPI 60
QY 61 EAGAINPENRIIPLPAKFWDCQRAAENRQGLTEYCTLSLPTKISRCPHLLDFFKV 120
DB 61 EAGEIHTENRVIPHLPAKFWDCQRAAENRQGLTEYCTLSLPTKISRCPHLLDFFKV 120
QY 121 RPDDLKLPDNTQKPEYLMKDGKSTATDITGPILQTYRAIANYEKTSSEWALSTG 180
DB 121 RPDDLKLPDNTQKPEYLMKDGKSTATDITGPILQTYRAIANYEKTSSEWALSTG 180
QY 181 DVVEVVEKSESGWFCOMKAKRGWIPASPLEDSDPEDEPNNYAGPYVAIKAYTAV 240
DB 181 DVVDVVEKSESGWFCOMKAKRGWIPASPLEDSDPEDEPNNYAGPYVAIKAYTAV 240
QY 241 EGDEVSLLEGEAVEVTHKLLDGW---KDDVTGYFPSPMYLQKSGQDVSOAQRQIK-RGAP 295
DB 241 EEDVSLSEGEAVEVTHKLLDGW---KDDVTGYFPSPMYLQKSGQDVSOAQRQIK-RGAP 295
QY 296 PRSSIRNAHSTHQSRKRLSODAYRRNSVRFLOQRRRQARPGPSGSPLEERQOTRS 355
DB 301 PRSTRIRNAQSIHQSRKRLSODTYRRNSVRFLOQRRRGRGPLSTDG-TKDNFSTPRV 359
QY 356 KQPAVPPRPSADLIINRCSESTKRLASAV 386
DB 360 KQPAVPPRPSDDLILHRCSTKRLKLTSAV 390
RESULT 8
Q99M65 PRELIMINARY; PRT; 276 AA.
ID Q99M65;
AC Q99M65;
DT 01-JUN-2001 (TReMBLrel. 17, Created)
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE ADAPTOR PROTEIN P47PHOX (FRAGMENT).
GN NCFL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;

```

[1]
RN SEQUENCE FROM N.A.
RA Gu Y., Xu Y., Souza R.F., Nwariaku F.E., Terada L.S.;
RT "Activation of c-Jun amino terminal kinase by a signaling oxidase.";
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF330625; AAK19516.1; -.
FT NON_TER 1
FT NON_TER 276 276
SQ SEQUENCE 276 AA; 31332 MW; 7FAF4BF05C89E220 CRC64;

Query Match 69.3%; Score 1411; DB 4; Length 276;
Best Local Similarity 97.8%; Pred. No. 1.4e-104;
Matches 270; Conservative 1; Mismatches 1; Indels 4; Gaps 1;

Qy 103 SLPTKISRCPLHLLDFKVRPDDKLPTDQNTKPKETYLMPKDGKSTATDITGPIILQYR 162
Db 1 SLPTKISRCPLHLLDFKVRPDDKLPTDQNTKPKETYLMPKDGKSTATDITGPIILQYR 60

Qy 163 AIANYEKTSGSEMALSTGDVVEVVEKESGWFCOMKAKRGWIPASFLPLDSPDETEDP 222
Db 61 AIADYEKTSGSEMALSTGDVVEVVEKESGWFCOMKTKRGWIPASFLPLDSPDETEDP 120

Qy 223 EPNVAGEPYVAIKAYTAVEGDEVSLLEGEAVEVIHKLLDGW----KDDVTGYFSPMYLQK 278
Db 121 EPNVAGEPYVAIKAYTAVEGDEVSLLEGEAVEVIHKLLDGWVIRKDDVTGYFSPMYLQK 180

Qy 279 SGQDVVSOAQROIKRGAPRRSSIRNAHSIRKRLSODAYRRNSVRFLOQRROARPG 338
Db 181 SGQDVVSOAQROIKRGAPRRSSIRNAHSIRKRLSODAYRRNSVRFLOQRROARPG 240

Qy 339 PQSPGSPLEERQTORSPQPAVPPRPSADLLNRC 374
Db 241 PQSPGSPLEERQTORSPQPAVPPRPSADLLNRC 276

RESULT 9
ID Q9JK56 PRELIMINARY; PRT; 310 AA.
AC Q9JK56;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE P47PHOX (FRAGMENT).
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN SEQUENCE FROM N.A.
RA Schieffer B., Luchtefeld M., Hilfiker A., Hilfiker-Kleiner D.,
RA Drexler H.;
RT "Role of NADPH oxidase in angiotensin II induced JAK/STAT
RT activation.";
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF260779; AAF70344.1; -.
DR InterPro; IPR001655; P47PHOX.
DR InterPro; IPR001683; PX.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00787; PX; 1.
DR Pfam; PF00018; SH3; 2.
DR PRINTS; PR00498; P47PHOX.
DR PRINTS; PR00452; SH3DOMAIN.
DR SMART; SM00312; PX; 1.
DR SMART; SM00326; SH3; 2.
DR PROSITE; PS50002; SH3; 2.
FT NON_TER 1
FT NON_TER 310 310
SQ SEQUENCE 310 AA; 35798 MW; A7B433F8186E2DBD CRC64;

Query Match 66.2%; Score 1346.5; DB 11; Length 310;
Best Local Similarity 80.6%; Pred. No. 2.2e-99;
Matches 248; Conservative 2; Mismatches 0; Indels 4; Gaps 1;

Qy 125 LKLPDNTQTKPKETYLMPKDGKSTATDITGPIILQYRAIANYEKTSGSEMALSTGDVVE 184
Db 1 LKLPDNTQTKPKETYLMPKDGKSTATDITGPIILQYRAIADYEKTSGSEMALSTGDVVE 60

Qy 185 VVEKESGWFCOMKAKRGWIPASFLPLDSPDETEDEPNVAGEPYVAIKAYTAVEGDE 244
Db 61 VVEKESGWFCOMKAKRGWIPASFLPLDSPDETEDEPNVAGEPYVAIKAYTAVEGDE 120

Qy 245 VSLLEGEAVEVIHKLLDGW----KDDVTGYFSPMYLQKSGQDVVSOAQROIKRGAPRRSS 300
Db 121 VSLLEGEAVEVIHKLLDGWVIRKDDVTGYFSPMYLQKSGQDVVSOAQROIKRGAPRRSS 180

Qy 301 IRNAHSIRKRLSODAYRRNSVRFLOQRROARPGSPGSPLEERQTORSPQPA 360
Db 181 IRNAHSIRKRLSODAYRRNSVRFLOQRROARPGSPGSPLEERQTORSPQPA 240
```

```

Matches 250; Conservative 27; Mismatches 28; Indels 5; Gaps 2;

Qy 21 SQHYVYMLVKQDLSEKVVYRRFTEIYEFHKTLEKEMPEIEAGAINPENRIIPHPAPKW 80
Db 1 SQHYVYMLVKQDLSEKVVYRRFTEIYEFHKTLEKEMPEIEAGIHTENRVRPHLPAPRW 60

Qy 81 FDGQRAENROGTLTEYCSTLMSLPTKISRCPLHLLDFKVRPDDKLPTDQNTKPKETYL 140
Db 61 YDQGRAESRQGLTEYFNLMGLPMKISRCPHLLNFEKVRPDDLKLPNDOSOVKKPETYL 120

Qy 141 MPKDGKSTATDITGPIILQYRAIANYEKTSGSEMALSTGDVVEVVEKESGWFCOMKA 200
Db 121 TAKDGKNVADIMGPIILQYRAIADYEKSGKTENTVATGDVVDVVEKESGWFCOMKP 180

Qy 201 KRGWIPASFLPLDSPDETEDEPNVAGEPYVAIKAYTAVEGDEVSLLEGEAVEVIHKLL 260
Db 181 NRGWIPASFLPLDSPDETEDEPNVAGEPYVAIKAYTAVEGDEVSLLEGEAVEVIHKLL 240

Qy 261 DGW----KDDVTGYFSPMYLQKSGQDVVSOAQROIK-RGAPRRSSIRNAHSIRKRL 315
Db 241 DGWVVRKGDITGYFSPMYLQKAGEKITQAOQRIKRGAPRRSSIRNAHSIRKRL 300

Qy 316 SQDAYRRNSV 325
Db 301 SQDYRRNSV 310

RESULT 10
Q9BX17 PRELIMINARY; PRT; 254 AA.
ID Q9BX17
AC Q9BX17;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DE ADAPTOR PROTEIN P47PHOX (FRAGMENT).
GN NCF1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN SEQUENCE FROM N.A.
RA TISSUP-UMBILICAL VEIN;
RA Gu Y., Xu Y., Souza R.F., Nwariaku F.E., Terada L.S.;
RT "Activation of c-Jun amino terminal kinase by a signaling oxidase.";
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF330626; AAK19517.1; -.
FT NON_TER 1
FT NON_TER 134 134
FT NON_TER 254 254
SQ SEQUENCE 254 AA; 28736 MW; FD7CC225ED7EDCF3 CRC64;
```

```
QY 361 VPPRPSADILNRC 374
Db 241 VPPRPSADILNRC 254

RESULT 11
ID 089032 PRELIMINARY; PRT; 1124 AA.
AC 089032;
DT 01-NOV-1998 (TReMBLrel. 08, Created)
DT 01-NOV-1998 (TReMBLrel. 08, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE FISH PROTEIN.
GN B416N2.2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Lock P., Abram C.L., Gibson T., Courtneidge S.A.;
RX MEDLINE=98353460; PubMed=9687503;
RT "A new method for isolating tyrosine kinase substrates used to
RT identify fish, an SH3 and PX domain-containing protein, and Src
RT substrate."
RL EMBO J. 17:4346-4357(1998).
DR EMBL: AJ007012; CAA07416.1;
DR MGD: MGI:1298393; Fish.
DR InterPro: IPR001683; PX.
DR InterPro: IPR001452; SH3.
DR Pfam: PF00787; PX; 1.
DR Pfam: PF00018; SH3; 5.
DR SMART: SM00312; PX; 1.
DR SMART: SM00326; SH3; 5.
DR PROSITE: PS00002; SH3; 5.
DR PROSITE: PS00002; SH3; 5.
SQ SEQUENCE 1124 AA; 124170 MW; 2A001B1D1CE98A67 CRC64;

Query Match 24.8%; Score 504.5; DB 11; Length 1124;
Best Local Similarity 27.5%; Pred. No. 1.4e-31;
Matches 123; Conservative 78; Mismatches 151; Indels 95; Gaps 10;

QY 6 IRHIALGFEKRFVPSQHYVVMFLYKWDLSKVVYRFTETIYEPHKLKEMFPIEAGAI 65
Db 6 VQDATVWDVEKRSPSKHYVIINVTWSDTSQTYRYRSKFEDLQMLDKFPIEGQK 65

QY 66 NPENRIIPLPAKWFQDQ---AAENRGTLTEYCSLMSLPTKISRCPHLLDFKVRP 122
Db 66 DPKQRIIPFLPGKILFRSHIRDAVKRLKPIDEYCRALVRLPPIHSQCDEVFRFEARP 125

QY 123 DDLKLP-TDNOTKKPETYLMPKDGKSTATDTG-----PILOTYRAIANYEKTSSEM 175
Db 126 EDVNPPEKEDYSGSKRSKSVLSSWAESPCKDVGTADTNAEPMLEQYVVVSNYKKQENSEL 185

QY 176 ALSTGDVVEVSESGWFWCOMKAKRGWIPASFLEPLDSDPDETPDPNNYAG----- 228
Db 186 SLQAGEVDVIEKESGWFWFVSTSEQGWVPATYLEAQNGTRDSDINTSKTGEVSKRRK 245

QY 229 -----EPYVAIKAYTAVEGDEVSLLEGEAVEVTHKLLDGV-- 263
Db 246 AHLRLDRRLWTGLGMVNRQHSREERYVTVPYTSQSKDEIGFEKGVTVVIRKKNLEGWY 305

QY 264 -----KDDVTGYFPSPMYLQSGODVSAQROI- 363
Db 306 IRYLKE---GWAPASYLKAKDDDLPTKKNLADGVEIIGNIMETSLNLLKASGDKEAP 362

QY 296 PR-----RSSIRNAHSIHQSRKLSQDAYRRNSVRFELQRRRQAR 336
Db 363 ABEGSEAPITKKEISLPILCNASNGSALAPTERTSKLAQGS--PAVARIAPQRAISS 420

QY 337 PQPQSGPGLSEERQTSKPKQPAVPP 363
Db 421 PNLTRPPPP--RESSLGFLPKPPEPP 446
```

```
RESULT 12
Q9H462 PRELIMINARY; PRT; 1031 AA.
ID Q9H462;
AC Q9H462;
DT 01-MAR-2001 (TReMBLrel. 16, Created)
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE B416N2.2 (SIMILAR TO MURINE FISH (AN SH3 AND PX DOMAIN-CONTAINING
DE PROTEIN, AND SRC SUBSTRATE)) (FRAGMENT).
GN B416N2.2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Chapman J.;
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL121929; CAC10184.1;
DR InterPro: IPR001452; SH3.
DR Pfam: PF00018; SH3; 5.
DR SMART: SM00326; SH3; 5.
DR PROSITE: PS00002; SH3; 5.
DR NON_TER 1
FT NON_TER 1
SQ SEQUENCE 1031 AA; 113127 MW; CF673CB96D2443E7 CRC64;

Query Match 16.4%; Score 334.5; DB 4; Length 1031;
Best Local Similarity 25.4%; Pred. No. 4.2e-18;
Matches 89; Conservative 58; Mismatches 110; Indels 93; Gaps 9;

QY 101 LMSLPTKISRCPHLLDFKVRPDDLKLP-TDNOTKKPETYLMPKDGKSTATDTG----- 154
Db 2 LVRLPPHISQCDEVFRFEARPEDVNPPEKEDYSGSKRSKSVLSSWAESPCKDVGTADATA 61

QY 155 -PILOTYRAIANYEKTSSEMALSTGDVVEVSESGWFWCOMKAKRGWIPASFLEPL 213
Db 62 EPMILEQYVVVSNYKKQENSELSQAGEVDVIEKESGWFWFVSTSEQGWVPATYLEAQ 121

QY 214 DSPDETEPEPNYAG-----EPYVAIKAYTAVEGD 243
Db 122 NGTRDSDSDINTSKTGEVSKRRKKAHLRLDRRLTGLGMVNRQHSREERYVTVPYTSQSKD 181

QY 244 EVSLLEGEAVEVTHKLLDGV-----KDDVTGYFPSPMYLQSGODVSAQROI----- 290
Db 182 ETGFEKGVTVVIRKKNLEGWVYRYLQKE---GWAPASYLKAKDDDLPTKKNLAGPVEI 238

QY 291 -----KRGAPPR-----RSSIRNAHSIHQSRK 313
Db 239 IGNIWEISNLNKKASGDKEPTPAEGEGHEAPIAKKEISLPILCNASNGSAGVGPDRTVS 298

QY 314 RLSQDAYRRNSVRFELQRRRQARPGPSPGSPLEEFERQTSKPKQPAVPP 363
Db 299 RLQAGS--PAVARIAPQRAISSPNLRTPPP--RESSLGFLPKPPEPP 345

RESULT 13
Q43302 PRELIMINARY; PRT; 940 AA.
ID Q43302;
AC Q43302;
DT 01-JUN-1998 (TReMBLrel. 06, Created)
DT 01-JUN-1998 (TReMBLrel. 06, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE KIAA0418.
GN KIAA0418.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
```


RC TISSUE-BRAIN;
RA Ishikawa K., Nagase T., Nakajima D., Seki N., Ohira M., Miyajima N.,
RA Tanaka A., Kotani H., Nomura N., Ohara O.;
RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB007878; BAA24848.1; -
DR InterPro; IPR001452; SH3.
DR Pfam; PF00018; SH3; 5.
DR PRINTS; PR00452; SH3DOMAIN.
DR PROSITE; PS00002; SH3; 5.
DR SMART; SM00326; SH3; 5.
SQ SEQUENCE 940 AA; 102617 MW; 1CA24D7BC03B8AC3 CRC64;

Query Match 13.5%; Score 275; DB 4; Length 940;
Best Local Similarity 26.9%; Pred. No. 2e-13;
Matches 70; Conservative 46; Mismatches 86; Indels 58; Gaps 7;

Qy 156 ILQTYRIANVETSGEMALSTGDDVVVVEKSESGWFWFQMKAKRGWIPASLEPLDS 215
Db 1 MLEQYVVVSNKKQENSELSQAGEVVDVTEKNSGWWFVSTSEEGWVPATVLEAQN 60
Qy 216 PDETEDEPNVAG--EPVAIKATAVGDEVSLEGEAVEVIHKLDDG-----KDD 266
Db 61 TRDSDINTSGTGEERKVTVPQYTSOSKDEIGFEKGVTVVEIRKNLDEGWVIRYLKRE- 119
Qy 267 VTGYFPSSMYLQSGDVSAQORQI-----KRGAPPR----- 297
Db 120 --GWAPASYLAKAKDDLLPTRKKNLGAHPVEIIGNIWEISNLLNKKASGDKETPPAEGEGE 177
Qy 298 -----RSSIRNAHSIHQRKRLSQDAYRRNSVRFLQRRRQARPGQSPG 343
Db 178 APIAKKEISLPILCNASGSAVGVPDRTVSLAQS--PAVARIAPQAQISSPNLRTP 235
Qy 344 SPLEEROTQRSKPOPAPVP 363
Db 236 PP-RRESSLGFOLPKPPEPP 254

RESULT 14
Q9GMZ5
ID Q9GMZ5 PRELIMINARY; PRT; 339 AA.
AC Q9GMZ5;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE P40-PHOX.
OS Tursiops truncatus (Atlantic bottlenose dolphin).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Cetacea; Odontoceti; Delphinidae;
OC Tursiops.
OX NCBI_TaxID=9739;
RN [1]
RP SEQUENCE FROM N.A.
RA Inoue Y., Itou T., Sakai T.;
RT "Molecular Cloning and Identification of Bottle-Nosed Dolphin p40-phox.";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB038267; BAB11805.1; -
DR InterPro; IPR000919; NCF_P40.
DR InterPro; IPR000108; Neu_cyt_fact_2.
DR InterPro; IPR000270; OPR.
DR InterPro; IPR001683; PX.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00564; OPR; 1.
DR Pfam; PF00787; PX; 1.
DR Pfam; PF00018; SH3; 1.
DR PRINTS; PR00497; P4OPHOX.
DR PRINTS; PR00499; P67PHOX.
DR PRINTS; PR00452; SH3DOMAIN.
DR SMART; SM00016; OPR; 1.
DR SMART; SM00312; PX; 1.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS00002; SH3; 1.

SQ SEQUENCE 339 AA; 39363 MW; B6306ADD3D44E686 CRC64;

Query Match 9.7%; Score 197; DB 6; Length 339;
Best Local Similarity 25.7%; Pred. No. 8.4e-08;
Matches 83; Conservative 48; Mismatches 110; Indels 82; Gaps 13;

Qy 8 HIALLGFEKRFVPSQHYVYMFVVKQDLSEKVVYRRFTEIYEFHKTLEKMPPIEAGAINP 67
Db 25 NIADIEKKGF--TSHFVFEVTKGSKYLIYRRYRQFYALQSKLEERF-----GP 75
Qy 68 ENR-----IIPHPAPKWDG--QRAAENROGTUTETCYSTLMSLPTKISRCPHLLD--- 116
Db 76 ENKASPYICILPTLPA-KVYVGVRQEIAEMRIPALNAYMKHLLSLPIWV-----LMDEDV 129
Qy 117 ---FEKVRPDLKLPDNDQTKPKETYLMPKDGKSTATDITPIILQTYRATANYEKTSGS 173
Db 130 RIFFYQSSYDAEQVPPQALRRURPT----RKVKSTSPQAGFDRMAAPRAEALFDFTGNS 185
Qy 174 EMALS--TGDVVEVVEKSESGWFWFQMKAKRGWIPASLEPLDSDPEDTDP----- 222
Db 186 KLELNFKYGDVIFLLSRINKDWLEGTVRGTTGIFPVSEVKILKDFPEEDPTNWLRCYY 245
Qy 223 -----EPNYAGEPY-----VAIKAYTAVEGDEVSLLEGEAVE 254
Db 246 EDTISTIKDIAVEEDLKSTPLFKDLLELMRREFQREDIALN-YQDAEGDLVRLLSDEVDV 304
Qy 255 VIHKLDDG-----WKDDVT 268
Db 305 LMVKWQTGLPSQKHLFPWKLHIT 327

RESULT 15
Q9NZM2
ID Q9NZM2 PRELIMINARY; PRT; 1248 AA.
AC Q9NZM2;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE INTERSECTIN 2 SHORT ISOFORM.
DE ITS2.
GN Homo sapiens (Human).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Pucharcos C., Estivill E., de la Luna S.;
RT "Cloning and characterization of human intersectin 2: a protein involved in endocytosis.";
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
CC -I- SIMILARITY: TO EF-HAND FAMILY.
DR EMBL; AF182199; AAF59904.1; -
DR InterPro; IPR002048; EF-hand.
DR InterPro; IPR000261; EPS15_repeat.
DR InterPro; IPR000108; Neu_cyt_fact_2.
DR InterPro; IPR001230; Prenyltn.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00036; ehand; 2.
DR Pfam; PF00018; SH3; 5.
DR PRINTS; PR00499; P67PHOX.
DR PRINTS; PR00452; SH3DOMAIN.
DR SMART; SM00054; EFh; 2.
DR SMART; SM00027; EH; 2.
DR SMART; SM00326; SH3; 5.
DR PROSITE; PS00018; EF_HAND; UNKNOWN_1.
DR PROSITE; PS00294; PRENYLATION; UNKNOWN_1.
DR PROSITE; PS00002; SH3; 5.
KW Calcium-binding.
SQ SEQUENCE 1248 AA; 141718 MW; A325BC186C76A97D CRC64;

Query Match 9.2%; Score 188; DB 4; Length 1248;

```
Best Local Similarity 22.1%; Pred. No. 2.5e-06;
Matches 60; Conservative 58; Mismatches 102; Indels 52; Gaps 9;

QY 64 AINPENRIIPHLPAKWFQDQRAAENROGTLTEYCSLMSLPTKISRCPHLLDFFKVRPD 123
Db 823 AVSPRKALLP-----PTVSLSATSTSSSEPLSSNQPASVTDYQNVSF 864
QY 124 DLKLPDNOTKKPEYLMFKDCKSTATDITGPI-----ILQTYR--AIANYEKTSGSEMA 176
Db 865 NLTVNTSWOKKSAFT-----RTVSPGSVSPHGGQGVVENLKAQALCSWTAKKDNHLN 917
QY 177 LSTGDVVEYVEKESGWWFCQMKARGWIPASFLEPLDSPD-ETEDPEFNVA----- 227
Db 918 FSKHDIITVLEQEN-WFGEVHGGRGWFPKSVYKIIPGSEVKREEPEALYAAVNKKPTS 976
QY 228 -----GEPYVAIKAYTAVEGDEVSLLEGEAVEVIHKLLDGWK---DDVTGYPPSMYLOKS 279
Db 977 AAYSVEEYIALPYSSVEPGDLTFTEGEIILVTQKDGEMWTGSGIDRSGIFPPSNYVKPK 1036
QY 280 GODVSOAQRIKRGAPPRSSIRNAHSIHQS 311
Db 1037 DOESFGSAS--KSGASNKKPEIAQVTSAYVAS 1066
```

Search completed: January 11, 2002, 09:11:05
Job time: 207 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 11, 2002, 09:07:58 ; Search time 13.31 Seconds
(without alignments)
1063.308 Million cell updates/sec

Title: US-09-820-005-2
Perfect score: 2035
Sequence: 1 MGDFTIRHALLGFEKRVFP.....ADLILNRCSESTKRLASAV 386

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	2018	99.2	390	1	NCF1_HUMAN
2	1754	86.2	392	1	NCF1_BOVIN
3	1689	83.0	390	1	NCF1_MOUSE
4	188.5	9.3	339	1	NCF4_MOUSE
5	172.5	8.5	339	1	NCF4_HUMAN
6	157	7.7	520	1	ITSN_HUMAN
7	136.5	6.7	671	1	VINE_HUMAN
8	135.5	6.7	733	1	VINE_MOUSE
9	133.5	6.6	1634	1	PK3B_HUMAN
10	133.5	6.6	2472	1	SPCN_HUMAN
11	131.5	6.5	303	1	CRKL_HUMAN
12	131.5	6.5	2415	1	SPCA_DROME
13	131.5	6.5	2477	1	SPCN_CHICK
14	129.5	6.4	303	1	CRKL_MOUSE
15	129.5	6.4	517	1	FGF_MOUSE
16	127.5	6.3	505	1	SRK1_SPOA
17	127.5	6.3	1149	1	DBS_MOUSE
18	127	6.2	1244	1	SLAL_YEAST
19	124	6.1	305	1	CRK_CHICK
20	122	6.0	1147	1	MYSB_ACACA
21	121.5	6.0	501	1	UVS2_NEUCR
22	121	5.9	536	1	SCD2_SCHPO
23	121	5.9	642	1	YB65_SCHPO
24	119.5	5.9	359	1	YKAT_CAEEL
25	119.5	5.9	403	1	STAC_MOUSE
26	119	5.8	646	1	Y142_HUMAN
27	116	5.7	1113	1	MYSD_DICDI
28	115	5.7	381	1	NCK2_HUMAN
29	114.5	5.6	506	1	SRK4_SPOA
30	113.5	5.6	695	1	AMPH_HUMAN
31	113.5	5.6	1333	1	CC25_CANAL
32	113	5.6	343	1	SNXG_HUMAN
33	113	5.6	377	1	NCK1_HUMAN

34	113	5.6	526	1	SRC_AVISR
35	113	5.6	532	1	SRC_CHICK
36	113	5.6	557	1	SRC_AVIS1
37	113	5.6	568	1	SRC_AVIS2
38	113	5.6	587	1	SRC_AVIS2
39	112	5.5	304	1	CRK_MOUSE
40	111.5	5.5	304	1	CRK_RAT
41	111.5	5.5	402	1	STAC_HUMAN
42	110	5.4	633	1	YHR4_YEAST
43	109.5	5.4	296	1	CRK_XENLA
44	108	5.3	304	1	CRK_HUMAN
45	107.5	5.3	861	1	SM4D_MOUSE

ALIGNMENTS

RESULT 1	
NCFL_HUMAN	STANDARD; PRT; 390 AA.
ID NCFL_HUMAN	AC P14598;
DT 01-APR-1990	(Rel. 14, Created)
DT 01-MAY-1992	(Rel. 22, Last sequence update)
DT 20-AUG-2001	(Rel. 40, Last annotation update)
DE NEUTROPHIL CYTOSOL FACTOR 1 (NCF-1) (NEUTROPHIL NADPH OXIDASE FACTOR	
DE 1) (47 KDA NEUTROPHIL OXIDASE FACTOR) (P47-PHOX) (NCF-47K) (47 KDA	
DE AUTOSOMAL CHRONIC GRANULOMATOUS DISEASE PROTEIN).	
GN NCF1.	
OS Homo sapiens (Human).	
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
OX NCBI_TaxID=9606;	
RN [1]	
RP SEQUENCE FROM N.A.	
RX MEDLINE=89332501; PubMed=2547247;	
RA Lomax K.J., Leto T.B., Nuno H., Gallin J.I., Malech H.L.;	
RT "Recombinant 47-kilodalton cytosol factor restores NADPH oxidase in	
RT chronic granulomatous disease.";	
RL Science 245:409-412(1989).	
RN [2]	
RP SEQUENCE FROM N.A.	
RX MEDLINE=89386707; PubMed=2550933;	
RA Volpp B.D., Nauseef W.M., Clark R.A.;	
RT "Cloning of the cDNA and functional expression of the 47-kilodalton	
RT cytosolic component of human neutrophil respiratory burst oxidase.";	
RL Proc. Natl. Acad. Sci. U.S.A. 86:7195-7199(1989).	
RN [3]	
RP REVISIONS.	
RA Volpp B.D., Nauseef W.M., Clark R.A.;	
RL Proc. Natl. Acad. Sci. U.S.A. 86:9563-9563(1989).	
RN [4]	
RP SEQUENCE FROM N.A.	
RX MEDLINE=90377229; PubMed=2398896;	
RA Rodaway A.R.F., Teahan C.G., Casimir C.M., Segal A.W., Bentley D.L.;	
RT "Characterization of the 47-kilodalton autosomal chronic	
RT granulomatous disease protein: tissue-specific expression and	
RT transcriptional control by retinoic acid.";	
RL Mol. Cell. Biol. 10:5388-5396(1990).	
CC -1- FUNCTION: NCF2, NCF1, AND A MEMBRANE BOUND CYTOCHROME B558 ARE	
CC REQUIRED FOR ACTIVATION OF THE LATENT NADPH OXIDASE (NECESSARY	
CC FOR SUPEROXIDE PRODUCTION).	
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.	
CC -1- DISEASE: NCF1 IS ABSENT IN MOST PATIENTS WITH AUTOSOMAL RECESSIVE	
CC CHRONIC GRANULOMATOUS DISEASE (AR-CGD).	
CC -1- SIMILARITY: CONTAINS 2 SH3 DOMAINS.	
CC -1- SIMILARITY: CONTAINS 1 PX DOMAIN.	
CC -1- DATABASE: NAME=NCFlbase; NOTE=NCFl deficiency database;	
CC WWW="http://www.uta.fi/inf/bioinfo/NCFlbase/"	

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

```
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M25665; AAA57209.1; -.
DR EMBL; M55067; AAA59901.1; -.
DR PIR; A39249; A39249.
DR MIM; 233700; -.
DR InterPro; IPR001655; P47PHOX.
DR InterPro; IPR001683; PX.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00787; PX; 1.
DR PRINTS; PR00452; SH3DOMAIN.
DR PRINTS; PR00498; P47PHOX.
DR SMART; SM00312; PX; 1.
DR SMART; SM00326; SH3; 2.
DR PROSITE; PS50002; SH3; 2.
KW SH3 domain; Chronic granulomatous disease.
FT DOMAIN 9 128
FT DOMAIN 211 254 ASP/GLU-RICH (HIGHLY ACIDIC).
FT DOMAIN 292 390 ARG/LYS-RICH (HIGHLY BASIC).
FT DOMAIN 156 215 SH3 1.
FT DOMAIN 226 285 SH3 2.
SQ SEQUENCE 390 AA; 44682 MW; B11B256516F3AA9B CRC64;

Query Match 99.2%; Score 2018; DB 1; Length 390;
Best Local Similarity 98.7%; Pred. No. 1.4e-140;
Matches 385; Conservative 1; Mismatches 0; Indels 4; Gaps 1;

QY 1 MGDTEFIRHALLGFEKRFVPSQHYVYMFVLKQDLSKRVYRRFTEIYEFHKLKEMFPI 60
DB 1 MGDTEFIRHALLGFEKRFVPSQHYVYMFVLKQDLSKRVYRRFTEIYEFHKLKEMFPI 60

QY 61 EAGAINPENRIIPHLPAKWFQDQRAAENRQGLTEYCYSTLMSLPTKISRCPLHLLDFKV 120
DB 61 EAGAINPENRIIPHLPAKWFQDQRAAENRQGLTEYCYSTLMSLPTKISRCPLHLLDFKV 120

QY 121 RPDDLKLPDNTQTKPEYLMKDKCKSTATDITGPIIIQTVRAIANYEKTSGSEMASTG 180
DB 121 RPDDLKLPDNTQTKPEYLMKDKCKSTATDITGPIIIQTVRAIANYEKTSGSEMASTG 180

QY 181 DVVEVVEKSEGWFCOMKAKRGWIPASFLPLDSPDETEPEPNYAGEPYVAIKAYTAV 240
DB 181 DVVEVVEKSEGWFCOMKAKRGWIPASFLPLDSPDETEPEPNYAGEPYVAIKAYTAV 240

QY 241 EGDEVSLLEGEAVEVIHKLKLGW----KDDVTGYPPSMYLOKSGQDVSAQRIKRGAPP 296
DB 241 EGDEVSLLEGEAVEVIHKLKLGW----KDDVTGYPPSMYLOKSGQDVSAQRIKRGAPP 296

QY 297 RRSIRNAHSHORSKRRLSQDAYRNRNVRFLQORRRQARPCQSPGSPLEERQTSK 356
DB 297 RRSIRNAHSHORSKRRLSQDAYRNRNVRFLQORRRQARPCQSPGSPLEERQTSK 356

QY 357 POPAVPPRPSADLIILNRCSESTKRLASAV 386
DB 357 POPAVPPRPSADLIILNRCSESTKRLASAV 386

RESULT 2
NCFL_BOVIN
ID NCFL_BOVIN STANDARD; PRT; 392 AA.
AC 077774;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE NEUTROPHIL CYTOSOL FACTOR 1 (NCF-1) (NEUTROPHIL NADPH OXIDASE FACTOR
DE 1) (47 KDA NEUTROPHIL OXIDASE FACTOR) (P47-PHOX) (NCF-47k).
GN NCFL.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;

Bovidae; Bovinae; Bos.
NCBI_taxID=9913;
[1]
SEQUENCE FROM N.A.
Bunger P.L., Siemsen D., Quinn M.T.;
"Cloning and expression of bovine NADPH oxidase proteins p47-phox and
p67-phox.";
Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
-!- FUNCTION: NCF2, NCF1, AND A MEMBRANE BOUND CYTOCHROME B558 ARE
REQUIRED FOR ACTIVATION OF THE LATENT NADPH OXIDASE (NECESSARY
FOR SUPEROXIDE PRODUCTION).
-!- SUBCELLULAR LOCATION: CYTOPLASMIC.
-!- SIMILARITY: CONTAINS 2 SH3 DOMAINS.
-!- SIMILARITY: CONTAINS 1 PX DOMAIN
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
```

```
Db 360 RPKQPAVPPRPSADLILHRCSESTKRKLASAV 392
RESULT 3
NCF1_MOUSE
ID NCF1_MOUSE STANDARD; PRT; 390 AA.
AC Q09014; Q07014; Q09J134;
DT 01-NOV-1995 (Rel. 32, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE NEUTROPHIL CYTOSOL FACTOR 1 (NCF-1) (NEUTROPHIL NADPH OXIDASE FACTOR
DE 1) (47 KDA NEUTROPHIL OXIDASE FACTOR) (P47-PHOX) (NCF-47K).
GN NCF1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Macrophage;
RX MEDLINE=94164697; PubMed=8119734;
RA Jackson S.H., Malech H.L., Kozak C.A., Lomax K.J., Gallin J.I.,
RA Holland S.M.;
RT "Cloning and functional expression of the mouse homologue of
RT p47phox.";
RL Immunogenetics 39:272-275(1994);
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Leukemia;
RX MEDLINE=98149672; PubMed=9490028;
RA Mizuki K., Kadomatsu K., Hata K., Ito T., Fan Q.-W., Kage Y.,
RA Fukumaki Y., Sakaki Y., Takeshige K., Sumimoto H.;
RT "Functional modules and expression of mouse p40(phox) and p67(phox),
RT SH3-domain-containing proteins involved in the phagocyte NADPH oxidase
RT complex.";
RL Eur. J. Biochem. 251:573-582(1998).
RN [3]
RP SEQUENCE FROM N.A.
RA Green E.D.;
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: NCF2, NCF1, AND A MEMBRANE BOUND CYTOCHROME B558 ARE
CC REQUIRED FOR ACTIVATION OF THE LATENT NADPH OXIDASE (NECESSARY
CC FOR SUPEROXIDE PRODUCTION).
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -!- SIMILARITY: CONTAINS 2 SH3 DOMAINS.
CC -!- SIMILARITY: CONTAINS 1 PX DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L11455; AAA50469.1; -
DR EMBL; AB002663; BAA25649.1; -
DR EMBL; AF267747; AAF90134.1; -
DR MGD; MGI:97283; Ncf1.
DR InterPro; IPR001655; P47PHOX.
DR InterPro; IPR001683; PX.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00787; PX; 1.
DR Pfam; PF00018; SH3; 2.
DR PRINTS; PR00452; SH3DOMAIN.
DR PRINTS; PR00498; P47PHOX.
DR SMART; SM00312; PX; 1.
DR SMART; SM00326; SH3; 2.
DR PROSITE; PS50002; SH3; 2.
KW SH3 domain.
FT DOMAIN 9 128 PX.
FT DOMAIN 211 254 ASP/GLU-RICH (HIGHLY ACTIDIC).
FT DOMAIN 292 390 ARG/LYS-RICH (HIGHLY BASIC).
```

```
FT DOMAIN 162 215 SH3 1.
FT DOMAIN 226 285 SH3 2.
FT CONFLICT 161 161 Y -> H (IN REF. 1).
FT CONFLICT 343 345 GOL -> RAA (IN REF. 1).
FT CONFLICT 344 344 Q -> P (IN REF. 3).
SQ SEQUENCE 390 AA; 44698 MW; A383DB953839CFGB CRC64;

Query Match 83.0%; Score 1689; DB 1; Length 390;
Best Local Similarity 80.8%; Pred. No. 1.7e-116;
Matches 316; Conservative 34; Mismatches 35; Indels 6; Gaps 3;

Qy 1 MGDTEIRHIALGLFEKRFVPSQHYVMFLVKWQDLSEKVVYRRFTEIYEFHKLKEMPEPI 60
Db 1 MGDTEIRHIALGLFEKRFVPSQHYVMFLVKWQDLSEKVVYRRFTEIYEFHKLKEMPEPI 60
Qy 61 EAGAINPENRIIPLPAPKWFEDGQRAAENROGTLTEYCSTLMSLPTKISRCPHLLDFPKV 120
Db 61 EAGEIHTENRVIPHLPAPRWFDGQRAAESROGTLTEYFNGLMGLPVKISRCPHLLDFPKV 120
Qy 121 RPDDLKLPDNTQTKKPEYLMKPKDCKGKSTATDITGPIIIQTYRAIANYEKTSSEMAISLG 180
Db 121 RPDDLKLPDNTQTKKPEYLMKPKDCKGKSTATDITGPIIIQTYRAIANYEKTSSEMAISLG 180
Qy 181 DVVEVVEKESGWFQCMKAKRGWIPASFLPLDSDPDEDEPEPNYAGEPYVAIKAYTAV 240
Db 181 DVVDVVEKESGWFQCMKAKRGWIPASFLPLDSDPDEDEPEPNYAGEPYVAIKAYTAV 240
Qy 241 EGDEVSLLEGEAVEVIHKLDDGW----KDDVTGYFPSPMYLQKSGQDVSAQRIK-RGAP 295
Db 241 EEDMSLSEGEAIEVIHKLDDGWVYVRKGDITGYSPMYLQKAGEITQAQRIKRGAP 300
Qy 296 PRSRIRNAHSIHQRSLKRLSDAYRRNSVRFLOQRRAQRCGPSPGSPLEERQOTORS 355
Db 301 PRSRIRNAQSIHQSRKRLSDQTYRRNSVRFLOQRRAQRCGPSPGSPLEERQOTORS 355
Qy 356 KPQPAVPPRPSADLILNRCSESTKRKLASAV 386
Db 360 KPQPAVPPRPSADLILHRCSESTKRKLTSAV 390

RESULT 4
NCF4_MOUSE
ID NCF4_MOUSE STANDARD; PRT; 339 AA.
AC P97369;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE NEUTROPHIL CYTOSOL FACTOR 4 (NCF-4) (NEUTROPHIL NADPH OXIDASE FACTOR
DE 4) (P40-PHOX) (P40PHOX).
GN NCF4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97152411; PubMed=8995189;
RA Zhan S., Kozak C.A., Zhan S., Chanock S.J.;
RT "Cloning and chromosomal localization of ncf4, the mouse homologue of
RT p40-phox.";
RL Immunogenetics 45:217-219(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Leukemia;
RX MEDLINE=98149672; PubMed=9490028;
RA Mizuki K., Kadomatsu K., Hata K., Ito T., Fan Q.-W., Kage Y.,
RA Fukumaki Y., Sakaki Y., Takeshige K., Sumimoto H.;
RT "Functional modules and expression of mouse p40(phox) and p67(phox),
RT SH3-domain-containing proteins involved in the phagocyte NADPH oxidase
RT complex.";
RL Eur. J. Biochem. 251:573-582(1998).
CC -!- FUNCTION: COMPONENT OF THE NADPH-OXIDASE, A MULTICOMPONENT ENZYME
```

```
CC SYSTEM RESPONSIBLE FOR THE OXYDATIVE BURST IN WHICH ELECTRONS ARE
CC TRANSPORTED FROM NADPH TO MOLECULAR OXYGEN, GENERATING REACTIVE
CC OXIDANT INTERMEDIATES. IT MAY BE IMPORTANT FOR THE ASSEMBLY AND/OR
CC ACTIVATION OF THE NADPH-OXYDASE COMPLEX.
CC -1- SUBUNIT: P40-PHOX ASSOCIATES PRIMARILY WITH P67-PHOX TO FORM A
CC COMPLEX WITH P47-PHOX.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 PX DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U59488; AAC53122.1; -.
DR EMBL; AB002665; BAA25651.1; -.
DR HSSP; P29355; 2SPM.
DR MGD; MGI:109186; Ncf4.
DR InterPro; IPR000919; NCF_P40.
DR InterPro; IPR000270; OPR.
DR InterPro; IPR001683; PX.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00564; OPR; 1.
DR Pfam; PF00787; PX; 1.
DR Pfam; PF00016; SH3; 1.
DR PRINTS; PR00452; SH3DOMAIN.
DR PRINTS; PR00497; P40PHOX.
DR SMART; SM00016; OPR; 1.
DR SMART; SM00312; PX; 1.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS00002; SH3; 1.
KW SH3 domain.
FT DOMAIN 24 143 PX.
FT DOMAIN 170 229 SH3.
SQ SEQUENCE 339 AA; 38734 MW; 62C4B3252D2C4B98 CRC64;

Query Match 9.3%; Score 188.5; DB 1; Length 339;
Best Local Similarity 26.2%; Pred. No. 8.2e-07;
Matches 80; Conservative 40; Mismatches 126; Indels 59; Gaps 10;

QY 15 EKRFPVSHYVMFLVKQDILSEKVVYRFTIYEFHTLKMFPDIEAGINPENRIIPH 74
DB 31 EKRGTFS-HEFVIEVTKGSKYLIYRRYQFYALQSKLEERFGPES-KNSPFTCNLPT 88
QY 75 LPAPKWFEG-QRAAENRQGTLTGYCSTLMSLPTKISRCPHLLDPFKVRPDDLKLPDNDQT 133
DB 89 LPAPKWFEG-QRAAENRQGTLTGYCSTLMSLPTKISRCPHLLDPFKVRPDDLKLPDNDQT 143
QY 134 KRPETLYMPKDGKSTATDITGPIL--LQTYRAIANYEKTSGSEMAIS--TGDVWEVWEVKS 189
DB 144 POALRRLRPRTRKIKGVSPQAGIMDRMEAPRAEALFDFTGNSKLSLSPKAGDVIFLLSKI 203
QY 190 ESWWFQCOMKAKRWIPASFLPLDSPDETDP-----EPNTYAG 228
DB 204 NKDWLEGTSGQATGIFPGSFVKILDFDEDDTNNWLRICYFYEDTGKTIKIDIAVEEDLS 263
QY 229 EPY-----VAIKATYAVEGVDSVLEGEAVEVHKLDDG-----W 263
DB 264 TPLFKDLALMRREFORREDIAL-SIQDAGDLVRLLSDEDVGLVWKQARGLPQSKRLFPW 322
QY 264 KDDVT 268
DB 323 KLHVT 327

RESULT 5
NCF4_HUMAN
>ID NCF4_HUMAN STANDARD; PRT; 339 AA.
```

```
AC Q15080; O60808; Q9NP45;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE NEUTROPHIL CYTOSOL FACTOR 4 (NCF-4) (NEUTROPHIL NADPH OXIDASE FACTOR
DE 4) (P40-PHOX) (P40PHOX).
GN NCF4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=94107216; PubMed=8280052;
RA Wientjes F.B., Hsuan J.J., Totty N.F., Segal A.W.;
RT "p40phox, a third cytosolic component of the activation complex of the
RT NADPH oxidase to contain src homology 3 domains.";
RL Biochem. J. 296:557-561(1993).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=96437051; PubMed=8839867;
RA Zhan S., Vazquez N., Zhan S., Wientjes F.B., Budarf M.L., Schrock E.,
RA Ried T., Green E.D., Chanock S.J.;
RT "Genomic structure, chromosomal localization, start of transcription,
RT and tissue expression of the human p40-phox, a new component of the
RT nicotinamide adenine dinucleotide phosphate-oxidase complex.";
RL Blood 88:2714-2721(1996).
RN [3]
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RX MEDLINE=99364495; PubMed=10437784;
RA Hasebe T., Someya A., Nagaoka I.;
RT "Identification of a splice variant mRNA of p40phox, an NADPH oxidase
RT component of phagocytes.";
RL FEBS Lett. 455:257-261(1999).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=20057165; PubMed=10591208;
RA Dunham I., Hunt A.R., Collins J.E., Bruskiewich R., Beare D.M.,
RA Clamp M., Smink L.J., Ainscough R., Almeida J.P., Babbage A.,
RA Bagguely C., Bailey J., Barlow K., Bates K.N., Beasley O., Bird C.P.,
RA Blakey S., Bridgeman A.M., Buck D., Burgess J., Burrill W.D.,
RA Burton J., Carder C., Carter N.P., Chen Y., Clark G., Clegg S.M.,
RA Cobley V., Cole C.G., Collier R.E., Connor R.E., Conroy D., Corby N.,
RA Coville G.J., Cox A.V., Davis J., Dawson E., Dhani P.D., Dockree C.,
RA Dodsworth S.J., Durbin R.M., Ellington A., Evans K.L., Fey J.M.,
RA Fleming K., French L., Garner A.A., Gilbert J.G.N., Goward M.B.,
RA Graham D., Griffiths M.N., Hall C., Hall R., Hall-Tamlyn G.,
RA Heathcote R.W., Ho S., Holmes S., Hunt S.E., Jones M.C., Kershaw J.,
RA Kimberley A., King A., Laird G.K., Langford C.F., Leversha M.A.,
RA Lloyd C., Lloyd D.M., Martyn I.D., Mashreghi-Mohammadi M.,
RA Mathews L., Mccann O.T., Mcclay J., McLaren S., Mcmurray A.A.,
RA Milne S.A., Mortimore B.J., Odell C.N., Pavitt R., Pearce A.V.,
RA Pearson D., Phillimore B.J., Phillips S.H., Plumb R.W., Ramsay H.,
RA Ramsey Y., Rogers L., Ross M.T., Scott C.E., Sehra H.K., Skuce C.D.,
RA Smalley S., Smith M.L., Soderlund C., Spragon L., Steward C.A.,
RA Sulston J.E., Swann R.M., Vaudin M., Wall M., Wallis J.M.,
RA Whiteley M.N., Willey D., Williams L., Williams S., Williamson H.,
RA Wilmer T.E., Wilming L., Wright C.L., Hubbard T., Bentley D.R.,
RA Beck S., Rogers J., Shimizu N., Minoshima S., Kawasaki K., Sasaki T.,
RA Asakawa S., Kudoh J., Shintani A., Shibuya K., Yoshizaki Y., Aoki N.,
RA Mitsuyama S., Roe B.A., Chen F., Chu L., Crabtree J., Deschamps S.,
RA Do A., Do T., Dorman A., Fang F., Fu Y., Hu P., Hua A., Kenton S.,
RA Lai H., Lao H.I., Lewis J., Lewis S., Lin S.-P., Loh P., Malaj E.,
RA Nguyen T., Pan H., Phan S., Qi S., Qian Y., Ray L., Ren Q., Shaull S.,
RA Sloan D., Song L., Wang Q., Wang Y., Wang Z., White J., Willingham D.,
RA Wu H., Yao Z., Zhan M., Zhang G., Chissoe S., Murray J., Miller N.,
RA Minx P., Fulton R., Johnson D., Bemis G., Bentley D., Bradshaw H.,
RA Bourne S., Cordes M., Du Z., Fulton L., Goela D., Graves T.,
RA Hawkins J., Hinds K., Kemp K., Latreille P., Layman D., Ozersky P.,
RA Rohlfing T., Scheet P., Walker C., Wamley A., Wohlmann P., Pepin K.,
RA Nelson J., Korf I., Bedell J.A., Hillier L., Mardis E., Waterston R.,
RA Wilson R., Emanuel B.S., Shaikh T., Kurahashi H., Saitta S.,
RA Budarf M.L., Mcdermid H.E., Johnson A., Wong A.C.C., Morrow B.E.,
```

RA Edelman L., Kim U.J., Shizuya H., Simon M.I., Dumanski J.P.,
 RA Peyrard M., Kedra D., Seroussi E., Fransson I., Tapia I., Bruder C.E.,
 RA O'Brien K.P., Wilkinson P., Bodenteich A., Hartman K., Hu X.,
 RA Khan A.S., Lane L., Tilahun Y., Wright H.,
 RT "The DNA sequence of human chromosome 22.",
 RL Nature 402:489-495(1999).
 CC -!- FUNCTION: COMPONENT OF THE NADPH-OXIDASE, A MULTICOMPONENT ENZYME
 CC SYSTEM RESPONSIBLE FOR THE OXYDATIVE BURST IN WHICH ELECTRONS ARE
 CC TRANSPORTED FROM NADPH TO MOLECULAR OXYGEN, GENERATING REACTIVE
 CC OXIDANT INTERMEDIATES. IT MAY BE IMPORTANT FOR THE ASSEMBLY AND/OR
 CC ACTIVATION OF THE NADPH-OXIDASE COMPLEX.
 CC -!- SUBUNIT: P40-PHOX ASSOCIATES PRIMARILY WITH P67-PHOX TO FORM A
 CC COMPLEX WITH P47-PHOX.
 CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
 CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; 1 (SHOWN HERE) AND 2; ARE
 CC PRODUCED BY ALTERNATIVE SPLICING.
 CC -!- TISSUE SPECIFICITY: EXPRESSION IS RESTRICTED TO HEMATOPOIETIC
 CC CELLS.
 CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
 CC -!- SIMILARITY: CONTAINS 1 PX DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U50729; AAB39970.1; -.
 DR EMBL; U50720; AAB39970.1; JOINED.
 DR EMBL; U50721; AAB39970.1; JOINED.
 DR EMBL; U50722; AAB39970.1; JOINED.
 DR EMBL; U50723; AAB39970.1; JOINED.
 DR EMBL; U50724; AAB39970.1; JOINED.
 DR EMBL; U50725; AAB39970.1; JOINED.
 DR EMBL; U50726; AAB39970.1; JOINED.
 DR EMBL; U50727; AAB39970.1; JOINED.
 DR EMBL; U50728; AAB39970.1; JOINED.
 DR EMBL; X77094; CAA54372.1; -.
 DR EMBL; AB025220; BAA89792.1; -.
 DR EMBL; AB025219; BAA89791.1; -.
 DR EMBL; AL008637; CAA15486.1; -.
 DR HSSP; P29355; 3SEM.
 DR MIM; 601488; -.
 DR InterPro; IPR000919; NCF_P40.
 DR InterPro; IPR000270; OPR.
 DR InterPro; IPR001683; PX.
 DR InterPro; IPR001452; SH3.
 DR Pfam; PF00564; OPR; 1.
 DR Pfam; PF00787; PX; 1.
 DR Pfam; PF00018; SH3; 1.
 DR PRINTS; PR00497; P40PHOX.
 DR SMART; SM00312; PX; 1.
 DR SMART; SM00016; OPR; 1.
 DR SMART; SM00326; SH3; 1.
 DR PROSITE; PS00002; SH3; 1.
 KW SH3 domain; Alternative splicing.
 FT DOMAIN 24 143 PX.
 FT DOMAIN 170 229 SH3.
 FT VARIANT 147 147 I -> L.
 FT /FTIDE=VAR_009314.
 FT /FTIDE=SSPPLKDLLELTRRREFOREDIALNVRDEGLD
 FT VRLLSDEVALMVQRAGLPQKRLFPWKLHITOKDNRYV
 FT NMP -> SVAWEGACPAFLPSLRPPLTSPSHGSLHSHK
 FT APSGQSHNAVTSRQPGWQPQHPFPPTPHFPQDASL
 FT LQVPTPLGTSRWKISALPY (IN ISOFORM 2).
 SQ SEQUENCE 339 AA; 39031 MW; 9A099AC0E834F90C CRC64;

Query Match 8.5%; Score 172.5; DB 1; Length 339;
 Best Local Similarity 24.4%; Pred. No. 1.2e-05;
 Matches 76; Conservative 48; Mismatches 116; Indels 71; Gaps 12;

QY 15 EKRFVPSOHVYVYMLVQWDLSEKVVYRRETEIYEFHKTLEKMFPIEAGAINPENR---- 70
 DB 31 EKRGFTS-HFVFIEVTKTGGSKYLIYRFRQFHALQSKLEERF-----GPDSSKSNL 82
 QY 71 --IIPHLPAKWFDC--QRAAENRQGTLETCSTLMSLPTKISRCPLHLDFEFKVRP-DDL 125
 DB 83 ACTLPTLPA-KVYGVKQIEAEMRIPALNAYMKSLSLSPVWVLMDEVDVRIFPYQSPYDSE 141
 QY 126 KLPTDNTQTKKPYLMPKDGKSTATDITGPILQIYRAIANYEKTSGSEMAILS--TGVV 183
 DB 142 QVQPAIRLRPRPT----RKVKSVPQGNVSVDRAAPAEALDFDTCKNSKLELNFAGDVI 197
 QY 184 EVVEKSESGWFWCOMKAKRGWIPASFLPLDSPDEDP----- 222
 DB 198 FLLSRINKDLGTVRGATGIFPLSFVKILKDFPEDDPTNWLRCYVYEDTISTIKDAV 257
 QY 223 EPNYAGEPY-----VAIKAYTAVGEVSLLEGEAVEVI-----H 257
 DB 258 EEDLSSTPLKDLLELTRRREFOREDIALN-YRDAEGDLVRLLSDEVALMVQRAGLPQ 316
 QY 258 KLLDGWKDDVT 268
 DB 317 KRLFPWKLHIT 327
 RESULT 6
 ITSN_HUMAN
 ID ITSN_HUMAN STANDARD; PRT; 520 AA.
 AC Q15811;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE INTERSECTIN (SH3 DOMAIN-CONTAINING PROTEIN SH3P17).
 GN ITSN OR SH3DIA.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Bone marrow;
 RX MEDLINE=98294438; PubMed=9630982;
 RA Sparks A.B., Hoffman N.G., McConnell S.J., Fowlkes D.M., Kay B.K.;
 RT "Cloning of ligand targets: systematic isolation of SH3 domain-
 RT containing proteins";
 RL Nat. Biotechnol. 14:741-744(1996).
 RN [2]
 RP GENE MAPPING.
 RX MEDLINE=98127038; PubMed=9465890;
 RA Chen H., Antonarakis S.E.;
 RT "The SH3DIA gene maps to human chromosome 21q22.1-->q22.2";
 RL Cytogenet. Cell Genet. 78:213-215(1997).
 CC -!- SIMILARITY: CONTAINS 4 SH3 DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; U61166; AAC50592.1; -.
 DR HSSP; P29354; 1GPD.
 DR MIM; 602442; -.
 DR InterPro; IPR001452; SH3.
 DR Pfam; PF00018; SH3; 4.
 DR PRINTS; PR00452; SH3DOMAIN.
 DR SMART; SM00326; SH3; 4.
 DR PROSITE; PS00002; SH3; 4.
 KW SH3 domain; Repeat.
 FT DOMAIN 116 177 SH3 1.

FT DOMAIN 284 342 SH3 2.
FT DOMAIN 374 438 SH3 3.
FT DOMAIN 455 514 SH3 4.
SQ SEQUENCE 520 AA; 57976 MW; 614043FIDB098C60 CRC64;

Query Match 7.7%; Score 157; DB 1; Length 520;
Best Local Similarity 28.8%; Pred. No. 0.00029;
Matches 47; Conservative 35; Mismatches 61; Indels 20; Gaps 9;

Qy 142 PKDGTATDITGPIIL--QYIRATANTKSGSEMASTGDVVEVBKSGWFCOMK 199
Db 358 PASLRKRVASPAKPVVSGEETIAQVIATATGPEQLTILAPGOLILIRKNPGWMEGELQ 417
Qy 200 A---KR--GWIPASFLPLDSP----DETEDEPNYAGE--PVVAIKAYTAVEGDEVS 247
Db 418 ARGKRQGWFPANVKKL--SPGTSKIPTPEPKSTALAAVCQVIGMVDYTAQNDELAF 476
Qy 248 LGEAVEVIHKL-LDGWKDDV---TGYFPPSMYLO-KSQGDVQSQ 285
Db 477 NKGQIINVLNKEDPDWKGVEYNGVGLFSPNVYKLTDMDFPSQ 519

RESULT 7

VINE_HUMAN
ID VINE_HUMAN STANDARD; PRT; 671 AA.
AC O60504; OSUQEA;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE VINEXIN (SH3-CONTAINING ADAPTOR MOLECULE-1) (SCAM-1).
GN SCAM1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., ALTERNATIVE SPLICING, AND MUTAGENESIS.
RC Tissue=Placenta;
RX MEDLINE=99102423; PubMed=9885244;
RA Kioka N., Sakata S., Kawauchi T., Amachi T., Akiyama S.K., Okazaki K.,
RA Yaen C., Yamada K.M., Aota S.-I.;
RT "Vinexin: a novel vinculin-binding protein with multiple SH3 domains
RT enhances actin cytoskeletal organization.";
RL J. Cell Biol. 144:59-69(1999).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM ALPHA).
RA Her J.-H., Gorman D., Miyajima A., Bolen J.B.;
RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP INTERACTION WITH SOS.
RX MEDLINE=20054482; PubMed=10585480;
RA Akamatsu M., Aota S.-I., Suwa A., Ueda K., Amachi T., Yamada K.M.,
RA Akiyama S.K., Kioka N.;
RT "Vinexin forms a signaling complex with Sos and modulates epidermal
RT growth factor-induced c-Jun N-terminal kinase/stress-activated
RT protein kinase activities.";
RL J. Biol. Chem. 274:35933-35937(1999).
CC -!- FUNCTION: VINEXIN ALPHA ISOFORM PROMOTES UPREGULATION OF ACTIN
CC STRESS FIBER FORMATION. VINEXIN BETA ISOFORM PLAYS A ROLE IN CELL
CC SPREADING AND ENHANCES THE ACTIVATION OF JNK/SAPK IN RESPONSE TO
CC EGF STIMULATION BY USING ITS THIRD SH3 DOMAIN.
CC -!- SUBUNIT: INTERACTS WITH VINCULIN BY THE FIRST TWO SH3 DOMAINS AND
CC THE PROLINE RICH REGION OF VINCULIN BINDS TO SOS (GUANINE
CC NUCLEOTIDE EXCHANGE FACTOR OF RAS AND RAC), THROUGH ITS THIRD SH3
CC DOMAIN. THE FORMATION OF THIS COMPLEX IS DOWN-REGULATED BY
CC PHOSPHORYLATION OF SOS.
CC -!- SUBCELLULAR LOCATION: BOTH ISOFORMS WERE LOCALIZED AT FOCAL
CC ADHESION AND CELL-CELL ADHESIONS SITES, VINEXIN BETA WAS ALSO
CC FOUND IN THE NUCLEUS.
CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS: ALPHA (SHOWN HERE) AND BETA; ARE
CC PRODUCED BY ALTERNATIVE SPLICING.
CC -!- TISSUE SPECIFICITY: BOTH ISOFORMS ARE EXPRESSED IN DIFFERENT

CC TISSUES LIKE HEART, PLACENTA, BRAIN, SKELETAL MUSCLE AND
CC PANCREAS. VINEXIN BETA ISOFORM IS ESPECIALLY FOUND IN LIVER.
CC -!- SIMILARITY: CONTAINS 3 SH3 DOMAINS.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; AF064807; AAD32304.1; -
DR EMBL; AF037261; AAC09244.1; -
DR HSSP; P07751; 1BK2.
DR InterPro; IPR000108; Neu_cyt_fact_2.
DR InterPro; IPR001452; SH3.
DR InterPro; IPR003127; Sorb.
DR Pfam; PF00018; SH3; 3.
DR Pfam; PF02208; Sorb; 1.
DR PRINTS; PR00452; SH3DOMAIN.
DR SMART; SM00326; SH3; 3.
DR SMART; SM00459; Sorb; 1.
DR PROSITE; PS50002; SH3; 3.
KW Cytoskeleton; SH3 domain; Repeat; Cell adhesion; Alternative splicing.
FT DOMAIN 117 164 SORB.
FT DOMAIN 380 439 SH3 1.
FT DOMAIN 454 515 SH3 2.
FT DOMAIN 612 671 SH3 3.
FT DOMAIN 380 515 BINDS TO VINCULIN.
FT DOMAIN 612 671 BINDS TO SOS.
FT VARSPIC 1 342 MISSING (IN ISOFORM BETA).
FT MUTAGEN 649 649 W->F: LOSS OF SOS-BINDING ABILITY.
FT MUTAGEN 667 667 Y->V: LOSS OF SOS-BINDING ABILITY.
FT CONFLICT 583 583 L -> F (IN REF. 1).
SQ SEQUENCE 671 AA; 75328 MW; 7D978711726BCD32 CRC64;

Query Match 6.7%; Score 136.5; DB 1; Length 671;
Best Local Similarity 22.6%; Pred. No. 0.013;
Matches 88; Conservative 51; Mismatches 140; Indels 11; Gaps 19;

Qy 46 EYIEFHKL-KEMFPIEAGAINPENRIIPHPAPKWFQDQRAAENR--QGTLYEYCSTLM 102
Db 278 ELAELSAELDKDLRAIETRLPSKSS-----PAPR-----RAPEORPPAGPASWSSVYP 327
Qy 103 SLP---TKISCPHLL-----DFFKVRPDDLKLPDNDQTKKPEYILMPKDGKST 148
Db 328 HAPYLGARSLSPHKMDGGSPFLGRDRF--VYPSSTRDPSASNGGSPARREKKRKA 385
Qy 149 ATDITGPIIQTIRAIANYKTSSEMALSTGDVVVEVVEKSGWFCOMKAKRWIPAS 208
Db 386 -----RLKDFDQAQSPKELTQKGDIVYTHKVDKNWLEGEHIGRLGIFPAN 432
Qy 209 FLEPLDSPDETD--EPNYAGEPVVAIKAYTAVGD---EVSLEGEAVEVIHKLDDGW 263
Db 433 YVEVLPA-DEIPKPIKPTYQVLEYGEAVQYTFKGDLEVELSPFKGHEICLRKVNENW 491
Qy 264 KDD-VT-----GYFPPSMYLOKSGQDVSOAQRIKRGAPRRSSIRNAHSIHQRSKRLSQ 317
Db 492 YEGRITGTGROGIFPASYVQ-----VSREPRLRCD 522
Qy 318 DAYRRNSVREFLQRRROARPGPQSGS-----PLEEER-OTQ 353
Db 523 DGPQLPTSPRTAAARSAR--HPSSPSALRSADPTDLGGQTSRPTGTFSPFTQEPRTQ 581
Qy 354 R-SKPQAPVPPR--PSADLIILNRCSESTKR 380
Db 582 NLGTGPPALSHSRGSPHPLDLGTSSPNTSQ 611

RESULT 8
VINE_MOUSE

DR Pfam: PF00435; spectrin; 23.
DR PRINTS: PR00452; SH3DOMAIN.
DR SMART: SM00054; EFh; 2.
DR SMART: SM00326; SH3; 1.
DR SMART: SM00150; SPEC; 20.
DR PROSITE: PS00018; EF_HAND; 1.
DR PROSITE: PS00002; SH3; 1.
KW Cytoskeleton; Membrane; Calmodulin-binding; Actin-binding;
KW Capping protein; Calcium-binding; Duplication; Repeat; SH3 domain.
FT DOMAIN 1 14
FT REPEAT 15 119
FT REPEAT 120 225
FT REPEAT 226 331
FT REPEAT 332 437
FT REPEAT 438 543
FT REPEAT 544 648
FT REPEAT 649 754
FT REPEAT 755 860
FT REPEAT 861 966
FT DOMAIN 967 1061
FT DOMAIN 967 1026
FT DOMAIN 1062 2251
FT REPEAT 1062 1167
FT REPEAT 1204 1309
FT REPEAT 1310 1415
FT REPEAT 1416 1521
FT REPEAT 1522 1628
FT REPEAT 1629 1734
FT REPEAT 1735 1840
FT REPEAT 1841 1946
FT REPEAT 1947 2053
FT REPEAT 2054 2166
FT REPEAT 2167 2251
FT DOMAIN 2252 2472
FT CA_BIND 2336 2347
FT CA_BIND 2379 2390
FT CONFLICT 1595 1595
SQ SEQUENCE 2472 AA; 284279 MW; CDA1964462A7028D CRC64;

Query Match 6.6%; Score 133.5; DB 1; Length 2472;
Best Local Similarity 19.3%; Pred. No. 0.12;
Matches 49; Conservative 45; Mismatches 91; Indels 69; Gaps 8;

QY 31 KQDLSEKVVYR-----FTEYEFHKTLEKMFIEAGAINPENRIIHLPA 78
DB 871 KWEALKASORQDLESQAQQVFADANEAEWSWREKEPIVGSTDYGKDE----- 922
QY 79 KWFQGRAA-----ENRQGTLTETCYSTLMSLPTKISRCPHLLDFFKVRPDDLKLPDNTQTK 134
DB 923 ---DSAEALLKHEALMSDLSAYGSSIOALREQAQSQRQV-----APTDDET 968
QY 135 KPEYLMKDGKSTATDTGPIILQTYRAIANEYKTSSEMALSTGDVVEVVEKSESQW 194
DB 969 K-----ELVLALY----DYEKSPREVTMKKGDIILTLNSTKQW 1005
QY 195 FQMKAKRGWIPASLEPLDSDPEDEPNYAGEPYVAIK-----AYTAVEGEVSL 247
DB 1006 KVEVNDROGFVPAIVKLLD-PAQSASRENLEEGSIALROEQIDNTRITKAGSVSL 1064
QY 248 LGEAVEVIHKLID 261
DB 1065 RMQVVEEYHLSLE 1078

RESULT 11
ID CRKL_HUMAN STANDARD; PRT; 303 AA.
AC P46109;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE CRK-LIKE PROTEIN.

GN CRKL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Splice;
RX MEDLINE=93368949; PubMed=8361759;
RA ten Hoeve J., Morris C., Heisterkamp N., Groffen J.;
RT "Isolation and chromosomal localization of CRKL, a human crk-like
gene.";
RL Oncogene 8:2469-2474(1993).
CC -!- FUNCTION: MAY MEDIATE THE TRANSDUCTION OF INTRACELLULAR SIGNALS.
CC -!- SIMILARITY: CONTAINS 1 SH2 DOMAIN.
CC -!- SIMILARITY: CONTAINS 2 SH3 DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X59656; CAA42199.1; -.
DR HSSP: Q64010; 1CKB.
DR MIN: 602007; -.
DR InterPro: IPR000980; SH2.
DR InterPro: IPR001452; SH3.
DR Pfam: PF00017; SH2; 1.
DR Pfam: PF00018; SH3; 1.
DR PRINTS: PR00401; SH2DOMAIN.
DR PRINTS: PR00452; SH3DOMAIN.
DR SMART: SM00252; SH2; 1.
DR SMART: SM00326; SH3; 2.
DR PROSITE: PS00001; SH2; 1.
DR PROSITE: PS00002; SH3; 1.
KW SH2 domain; SH3 domain.
FT DOMAIN 14 102
FT DOMAIN 123 183
FT DOMAIN 235 296
SQ SEQUENCE 303 AA; 33777 MW; 294CFLEE2CD44B81 CRC64;

Query Match 6.5%; Score 131.5; DB 1; Length 303;
Best Local Similarity 21.6%; Pred. No. 0.01;
Matches 50; Conservative 36; Mismatches 85; Indels 61; Gaps 9;

QY 103 SLPTK-----ISRCPHLLDFFKVRPDDLKLPDNTQTKPETYLMKDGKSTADI-T 153
DB 65 SLPNRRFKIGDOEFDHLPALLEFYKIH-----YLDTTTLIEPAPRYPSPMGSVSAPNLT 120
QY 154 GPILQTYRAIANEYKTSSEMALSTGDVVEVVEKSESQWFCQMKAKR-GWIPASFLEP 212
DB 121 AEDNLEYVRTLVDYFGNDAEDLPFKKGEILVIEKPEEQWWSARNKDGVRGMPVYVEK 180
QY 213 L--DSP-----DETEDEPNYAGEPYVAIKAYTAVEGEV-- 245
DB 181 LVRRSPHGKHGNNRNSNYGIPEPAHAYAQPTTTLPAVSGSPGAAITPLPSTQNGPVFA 240
QY 246 -----SILEGEAVEVIHKLIDG-WKDDVT---GVFPMSYLQ 277
DB 241 KAIQKRVPCAYDKTALALEVGDIVKVTNRNINGQWEGEVNGRKGLEFPPTHVK 292

RESULT 12
ID SPCA_DROME STANDARD; PRT; 2415 AA.
AC P13395; Q9W085; Q26340;
DT 01-JAN-1990 (Rel. 13, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)


```
FT REPEAT 1044 1151 SPECTRIN 11.
FT REPEAT 1152 1257 SPECTRIN 12.
FT REPEAT 1258 1363 SPECTRIN 13.
FT REPEAT 1364 1469 SPECTRIN 14.
FT REPEAT 1470 1576 SPECTRIN 15.
FT REPEAT 1577 1682 SPECTRIN 16.
FT REPEAT 1683 1788 SPECTRIN 17.
FT REPEAT 1789 1894 SPECTRIN 18.
FT REPEAT 1895 2001 SPECTRIN 19.
FT REPEAT 2002 2115 SPECTRIN 20.
FT REPEAT 2116 2229 SPECTRIN 21.
FT REPEAT 2230 2335 SPECTRIN 22.
FT REPEAT 2336 2441 SPECTRIN 23.
FT CA_BIND 2278 2389 EF-HAND 1 (POTENTIAL).
FT CA_BIND 2321 2332 EF-HAND 2 (POTENTIAL).
FT DOMAIN 970 1029 SH3.
FT CONFLICT 110 110 Q -> D (IN REF. 3).
FT CONFLICT 1668 1668 Q -> R (IN REF. 1).
SQ SEQUENCE 2415 AA; 278301 MW; F1F7F9B990EB0A37 CRC64;

Query Match 6.5%; Score 131.5; DB 1; Length 2415;
Best Local Similarity 15.3%; Pred. No. 0.16; Indels 133; Gaps 10;
Matches 56; Conservative 62; Mismatches 114;

Qy 31 KWQDLSEKVVYRR-----FTIEYEFHKLKEMFPIEAGAIENPENRIIPLHPAP 78
Db 874 QWNTLKEKSSQKQDLDLSLAHQYFADANEASWMEKEPIATGSDYKDE----- 925

Qy 79 KWFDCQRAENRQGLTE-----YCTLSMLPTKISRCPHLLDFFKVRPDDUKLPTDNTK 134
Db 926 ---DSSEALLKHEALVSDLEAFNGNTIQAQEQAKNC----- 959

Qy 135 KPETYLMPKDGKSTATDTPILITYRAIANEYKTSGEWALSTGDEVVEVVEKSESGW 194
Db 960 -----RQQTVPVDITGK---ECVVALDYTKSPREVSMMKKGDVITLNSNKKQW 1008

Qy 195 FQMKAKRGWIPASLEPLDS-----PDTEDEPEPNVAGPY----- 231
Db 1009 KVEVNDRGFGVPAAYIKKIDAGLSAQSNLVNHSIAKRNQINSQYDNLALARERQNK 1068

Qy 232 --VAIKATAVE-----GDEVSLLEGEAVEVVIKLLDGWKDDV 267
Db 1069 LNETVKAYVLVREAADLAQWIRKDNHQAQIADVVGEDL-----EEVEVLQKFFDNDL 1123

Qy 268 TG-----YPSMWLYKSGQDVSOAQROIKRGAPPRSSIRNAHSI 307
Db 1124 KANEVRLANMNEIAVOLTSLGQTEALKIQTQMQLNEKNWNLNQITAEKASQLGSAHEV 1183

Qy 308 HQRSR 312
Db 1184 QRFHR 1188

RESULT 13
ID SPCN_CHICK STANDARD; PRT; 2477 AA.
AC P07751.
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE SPECTRIN ALPHA CHAIN, BRAIN (SPECTRIN, NON-ERYTHROID ALPHA CHAIN)
DE (FODRIN ALPHA CHAIN).
GN SPTAN1 OR SPTA2.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89093238; PubMed=2910879;
RA Wasenius V.-M., Saraste M., Salven P., Eraama M., Holm M.,
RA Lehto V.-P.;
```

```
RT "Primary structure of the brain alpha-spectrin.";
RL J. Cell Biol. 108:79-93(1989).
RN [2]
RP REVISIONS.
RA Wasenius V.-M., Saraste M., Salven P., Eraama M., Holm L.,
RA Lehto V.-P.;
RL J. Cell Biol. 108:1177-1178(1989).
RN [3]
RP SEQUENCE OF 1695-2153 FROM N.A.
MEDLINE=85284928; PubMed=4029118;
RA Wasenius V.-M., Saraste M., Knowles J., Virtanen I., Lehto V.-P.;
RT "Sequencing of the chicken non-erythroid spectrin cDNA reveals an
internal repetitive structure homologous to the human erythrocyte
spectrin.";
RL EMBO J. 4:1425-1430(1985).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS) OF 965-1025.
MEDLINE=93063299; PubMed=1279434;
RA Musacchio A., Noble M., Paupit R., Wierenga R., Saraste M.;
RT "Crystal structure of a Src-homology 3 (SH3) domain.";
RL Nature 359:851-855(1992).
RN [5]
RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 969-1025.
MEDLINE=98363217; PubMed=9699637;
RA Martinez J.C., Pisabarro M.T., Serrano L.;
RT "Obligatory steps in protein folding and the conformational diversity
of the transition state.";
RL Nat. Struct. Biol. 5:721-729(1998).
RN [6]
RP STRUCTURE BY NMR OF 1763-1872.
MEDLINE=98022917; PubMed=9356261;
RA Pascual J., Pfuhl M., Walther D., Saraste M., Nilges M.;
RT "Solution structure of the spectrin repeat: a left-handed
antiparallel triple-helical coiled-coil.";
RL J. Mol. Biol. 273:740-751(1997).
RN [7]
RP STRUCTURE BY NMR OF 2320-2403.
MEDLINE=96067121; PubMed=7588621;
RA Trave G., Lacombe J.-P., Pfuhl M., Saraste M., Pastore A.;
RT "Molecular mechanism of the calcium-induced conformational change in
the spectrin EF-hands.";
RL EMBO J. 14:4922-4931(1995).
CC -!- FUNCTION: MORPHOLOGICALLY, SPECTRIN-LIKE PROTEINS APPEAR TO BE
RELATED TO SPECTRIN, SHOWING A FLEXIBLE ROD-LIKE STRUCTURE.
THEY CAN BIND ACTIN BUT SEEM TO DIFFER IN THEIR CALMODULIN-
BINDING ACTIVITY. IN NONERYTHROID TISSUES, SPECTRINS. IN
ASSOCIATION WITH SOME OTHER PROTEINS, MAY PLAY AN IMPORTANT
ROLE IN MEMBRANE ORGANIZATION.
CC -!- SUBUNIT: LIKE ERYTHROCYTE SPECTRIN, THE SPECTRIN-LIKE PROTEINS
ARE CAPABLE TO FORM DIMERS WHICH CAN FURTHER ASSOCIATE TO
TETRAMERS.
CC -!- DOMAIN: SPECTRIN-LIKE PROTEINS HAVE FIVE DOMAINS:
(1) N-TERMINAL DOMAIN (N),
(2) DOMAIN BETWEEN THE N-TERMINAL AND MIDDLE DOMAIN (NM),
(3) MIDDLE DOMAIN (M),
(4) DOMAIN BETWEEN THE MIDDLE AND C-TERMINAL DOMAIN (MC),
(5) C-TERMINAL DOMAIN (C).
CC NM AND MC DOMAINS ARE COMPOSED OF TYPICAL SPECTRIN 106 RESIDUES
REPEATS (1-8 FOR NM AND 12-19 FOR MC) AND ARE HOMOLOGOUS TO EACH
OTHER. N, M, AND C DOMAINS ARE COMPOSED OF SEQUENCES THAT DO NOT
FORM TYPICAL SPECTRIN REPEATS.
CC -!- SIMILARITY: BELONGS TO THE SPECTRIN FAMILY.
-!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X14518; CAA32662.1; -;
```

```
DR EMBL; X14519; CAA32663.1; ALT_SEQ.
DR EMBL; X02593; CAB51571.1; ALT_SEQ.
DR PIR; A30122; SJCHA.
DR PDB; 1SHG; 31-OCT-93.
DR PDB; 1AEY; 15-MAY-97.
DR PDB; 1AJ3; 07-JUL-97.
DR PDB; 1TUC; 01-AUG-96.
DR PDB; 1TUD; 01-AUG-96.
DR PDB; 1BK2; 16-FEB-99.
DR InterPro; IPR002048; EF-hand.
DR InterPro; IPR001452; SH3.
DR InterPro; IPR002017; Spectrin.
DR Pfam; PF00036; ehand; 2.
DR Pfam; PF00018; SH3; 1.
DR Pfam; PF00435; spectrin; 23.
DR PRINTS; PR00452; SH3DOMAIN.
DR SMART; SM00054; EFh; 2.
DR SMART; SM00326; SH3; 1.
DR SMART; SM00150; SPEC; 20.
DR PROSITE; PS00018; EF_HAND; 2.
DR PROSITE; PS00002; SH3; 1.
KW Cytoskeleton; Membrane; Calmodulin-binding; Actin-binding;
KW Capping protein; Calcium-binding; Duplication; Repeat; SH3 domain;
KW 3D-structure.
FT DOMAIN 1 14 N-TERMINAL DOMAIN.
FT REPEAT 15 119 1.
FT REPEAT 120 225 2.
FT REPEAT 226 331 3.
FT REPEAT 332 437 4.
FT REPEAT 438 543 5.
FT REPEAT 544 648 6.
FT REPEAT 649 754 7.
FT REPEAT 755 860 8.
FT REPEAT 861 966 9.
FT DOMAIN 967 1061 MIDDLE DOMAIN.
FT DOMAIN 967 1026 SH3.
FT DOMAIN 1062 2256 MC DOMAIN.
FT REPEAT 1062 1167 11.
FT REPEAT 1204 1309 12.
FT REPEAT 1310 1415 13.
FT REPEAT 1416 1521 14.
FT REPEAT 1522 1633 15.
FT REPEAT 1634 1749 16.
FT REPEAT 1740 1845 17.
FT REPEAT 1846 1951 18.
FT REPEAT 1952 2058 19.
FT REPEAT 2059 2171 20.
FT REPEAT 2172 2256 21.
FT DOMAIN 2257 2477 C-TERMINAL DOMAIN.
FT CA_BIND 2341 2352 EF-HAND 1 (POTENTIAL).
FT CA_BIND 2384 2395 EF-HAND 2 (POTENTIAL).
FT STRAND 371 374
FT STRAND 978 978
FT TURN 983 984
FT STRAND 985 985
FT STRAND 988 988
FT TURN 990 991
FT STRAND 993 998
FT STRAND 1004 1009
FT TURN 1010 1011
FT STRAND 1012 1017
FT HELIX 1018 1020
FT STRAND 1021 1024
SQ SEQUENCE 2477 AA; AD4C876994E6AB39 CRC64;

Query Match 6.5%; Score 131.5; DB 1; Length 2477;
Best Local Similarity 19.3%; Pred. No. 0.16;
Matches 49; Conservative 44; Mismatches 92; Indels 69; Gaps 8;

QY 31 KWQDLSEKVVYR-----FTEIYEPHKYTKEMFPTEAGAINPENRIPLPAP 78
QD 871 KWDSLKASQRRODLSLOAQOYFADANEQAQSWREKEPIVGSTDYGKDE----- 922
```

```
QY 79 KWFQDGRAA---ENRQGTLEYCSTLMSLPTKISRCPHLLDFKVRPDDLKLPDNDQTK 134
DB 923 ---DSAEALLKKHEALMSDL$AYGSSIOALRQAQSCROQV-----APDDETG 968
QY 135 KPETYLMFPGDKSTADITGPILQIOTYRAIANYEKTSSEMASTGDDVVVVVEKSESGW 194
DB 969 K-----ELVLALY---DYQEKSPREVTMKKGDIITLLNSINKDWW 1005
QY 195 FCQMAKRGWIPASFLPDLSPDETEDEPNVAGPYVAIK-----AYTAVEGDEVSL 247
DB 1006 KVEVNDROGFVPAAYVKKID-PAQSASRENLLLEQGSIALRQEQIDNQTLITKEVGSVSL 1064
QY 248 LGEAVEVEVIHKLLD 261
DB 1065 RMKQVEELYHSLLE 1078

RESULT 14
CRKL_MOUSE
ID CRKL_MOUSE STANDARD; PRT; 303 AA.
AC P47941;
DC 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE CRK-LIKE PROTEIN.
GN CRKL OR CRKOL.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL X CHA; TISSUE=Placenta;
RX MEDLINE=96038874; PubMed=7478571;
RA de Jong R.L., Haataja L., Voncken J.W., Heisterkamp N., Groffen J.;
RT "Tyrosine phosphorylation of murine Crkl.";
RL Oncogene 11:1469-1474(1995).
CC -!- FUNCTION: MAY MEDIATE THE TRANSDUCTION OF INTRACELLULAR SIGNALS.
CC -!- PTM: PHOSPHORYLATED ON TYROSINE. PHOSPHORYLATION IS PROMINENT
CC DURING EARLY DEVELOPMENT, BUT DECREASES AT LATER EMBRYONIC STAGES
CC AND IN NEWBORN MICE.
CC -!- SIMILARITY: CONTAINS 1 SH2 DOMAIN.
CC -!- SIMILARITY: CONTAINS 2 SH3 DOMAINS.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
DR EMBL; X90648; CAA62220.1;
DR HSSP; Q64010; 1CKB.
DR MGD; MGI:104686; Crkol.
DR InterPro; IPR000980; SH2.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00017; SH2; 1.
DR Pfam; PF00018; SH3; 1.
DR PRINTS; PR00401; SH2DOMAIN.
DR PRINTS; PR00452; SH3DOMAIN.
DR SMART; SM00252; SH2; 1.
DR SMART; SM00326; SH3; 2.
DR PROSITE; PS50001; SH2; 1.
DR PROSITE; PS50002; SH3; 1.
KW SH2 domain; SH3 domain; Phosphorylation.
FT DOMAIN 14 102 SH2
FT DOMAIN 123 183 SH3 1.
FT DOMAIN 235 296 SH3 2.
SQ SEQUENCE 303 AA; 33817 MW; B8126DDFF2053573 CRC64;
```

```
Query Match 6.4%; Score 129.5; DB 1; Length 303;
Best Local Similarity 21.6%; Pred. No. 0.015;
Matches 50; Conservative 35; Mismatches 86; Indels 61; Gaps 9;

QY 103 SLPTK-----ISRCPLLDFKVRPDDLKLPDNTQTKPKETVLMKPKDGKSTATDI-T 153
DB 65 SLPNRFKIGDQDFHLFALLEFYKH-----YLDTTLIEPAPRYPPSPVGSVAPNLP 120
QY 154 GPILIQTVRAIANYEKTSSESMALSTGDVVEVEKESGWFCEOMKAKR-GWIPASEFLEP 212
DB 121 AEENLEYVRLTYDFPCNDADLDFPKGELLVIEKPEQWWSARKDGRVGMIPVYVEK 180
QY 213 L--DSP-----DETEDEPNVAGPPYVAIKAYTAVEGDEV-- 245
DB 181 LVRSSPHGKHGRNNSYGIPEPAHAYAQPTTLPVASTPGAALNPLPSTQNGPVFA 240
QY 246 -----SLEGEAVEVIHKLDG-WKDDVT---GYFPPSYLQ 277
DB 241 KAIQKRVPCAYDKTALALEVGDIVKTRMNINGOWEGVNGRKLFPETHVK 292

RESULT 15
FGR_MOUSE
ID FGR_MOUSE STANDARD; PRT; 517 AA.
AC P14234; Q61404;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PROTO-ONCOGENE TYROSINE-PROTEIN KINASE FGR (EC 2.7.1.112) (P55-FGR)
DE (C-FGR).
GN Mus musculus (Mouse).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DBA/2J;
RX MEDLINE=89385605; PubMed=2674853;
RA Yi T.L., Willman C.L.;
RT "Cloning of the murine c-fgr proto-oncogene cDNA and induction of
RT c-fgr expression by proliferation and activation factors in normal
RT bone marrow-derived monocytic cells.";
RL Oncogene 4:1081-1087(1989).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/C; TISSUE=Monocytic leukemia;
RX MEDLINE=90191719; PubMed=2179817;
RA King F.J., Cole M.D.;
RT "Molecular cloning and sequencing of the murine c-fgr gene.";
RL Oncogene 5:337-344(1990).
CC -!- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE = ADP +
CC PROTEIN TYROSINE PHOSPHATE.
CC -!- SIMILARITY: CONTAINS 1 SH2 DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC -!- SIMILARITY: TO OTHER PROTEIN-TYROSINE KINASES IN THE CATALYTIC
CC DOMAIN BELONGS TO THE SRC SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X16440; CAA34463.1; -
CC EMBL; X52191; CAA36437.1; -
CC PIR; A33127; A33127.
CC PIR; S10072; S10072.
CC HSP; P00523; 2PTK.
CC MGD; MGI:95527; Fgr.
CC InterPro; IPR000719; Euk_pkinase.
```

```
DR InterPro; IPR000980; SH2.
DR InterPro; IPR001452; SH3.
DR InterPro; IPR001245; Tyr_kin.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF00017; SH2; 1.
DR Pfam; PF00018; SH3; 1.
DR PRINTS; PR00109; TYRKINASE.
DR PRINTS; PR00401; SH2DOMAIN.
DR PRINTS; PR00452; SH3DOMAIN.
DR SMART; SM00252; SH2; 1.
DR SMART; SM00326; SH3; 1.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00001; SH2; 1.
DR PROSITE; PS00002; SH3; 1.
KW Transferase; Tyrosine-protein kinase; Proto-oncogene; ATP-binding;
KW Phosphorylation; SH2 domain; SH3 domain.
FT DOMAIN 65 126 SH3.
FT DOMAIN 132 229 SH2.
FT DOMAIN 251 504 PROTEIN KINASE.
FT NP_BIND 257 265 ATP (BY SIMILARITY).
FT BINDING 279 279 ATP (BY SIMILARITY).
FT ACT_SITE 370 370 BY SIMILARITY.
FT MOD_RES 400 400 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT CONFLICT 41 41 N -> T (IN REF. 2).
FT CONFLICT 212 212 Q -> R (IN REF. 2).
SQ SEQUENCE 517 AA; 58867 MW; F655BDB4510F3076 CRC64;

Query Match 6.4%; Score 129.5; DB 1; Length 517;
Best Local Similarity 21.0%; Pred. No. 0.029;
Matches 65; Conservative 50; Mismatches 92; Indels 103; Gaps 16;

QY 119 KVRP---DDLKLPD--NOTKKPETYLMKPKDGKSTAT--DITGPILQI----- 160
DB 8 KLEPASKEDVGLGDFRSQTAERYYPDPTQGRNSSVFPQPTSPAFNLNTGNMRSISGTGV 67
QY 161 --YRAIANYEKTSSESMALSTGDVVEVEKESGWFCEQ--MKAKRGWIPASEFLEPLDSP 216
DB 68 TIFVALDYEARTEGDDLTFTKGEKPHILNNTYDWMWEARSLSSGHRGYVPSNYVAPVDSI 127
QY 217 DETE-----DPEPNY--AGEPY-----VAIKAYTAVEGDEVSLLE 249
DB 128 QAEEWFGKISRKAERQLLSGNPGCAFIRESETTKGAYSLIRDWQNRCDHIKHYK 187
QY 250 GEAVEVIHKLDGWKDDVTGYFPSPMYLQ-KSGQDVSAQQRQIKRG-----APPRSSIR 302
DB 188 -----IRKL-----DTGGYVITTTTRAQFDSIQDLVQHYMEVNDGLCYLLTAPCTTTKPQ 235
QY 303 NAHSIHQSRKRLSQDAYR--RNSVRFLOQRR-----ROARPG 338
DB 236 TL-----GLAKDAWEIDRNSIAL--ERRLGTGCGFDVWLGTWNCSTKVAVKTLKPG 284
QY 339 POSGSPLEEE 348
DB 285 TMSPRAFLEE 294
```

Search completed: January 11, 2002, 09:11:26
Job time: 208 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 11, 2002, 09:05:54 ; Search time 15.91 seconds
(without alignments)
1848.107 Million cell updates/sec

Title: US-09-820-005-2
Perfect score: 2035
Sequence: 1 MGDFTIRHIALGFKRFVP.....ADLILNRCSESTRKRLASAV 386

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_68.*

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2018	99.2	390	1 A39249	neutrophil cytosol
2	1666	81.9	388	2 I54525	leukemia-related p
3	275	13.5	940	2 T00056	hypothetical prote
4	179.5	8.8	1270	2 T09194	adaptor protein in
5	169.5	8.3	339	2 S39768	neutrophil cytosol
6	164.5	8.1	639	2 T13151	adapter protein CM
7	164	8.1	1011	2 T13055	dynamain associated
8	161	7.9	1094	2 T13053	dynamain associated
9	150.5	7.4	665	2 JC7191	85K c-Cbl-interact
10	147.5	7.2	290	2 T42526	hypothetical prote
11	145	7.1	1119	2 T50995	related to cytoske
12	142.5	7.0	1097	2 T31504	hypothetical prote
13	139	6.8	443	2 T27877	hypothetical prote
14	136	6.7	50	4 A58224	neutrophil cytosol
15	135.5	6.7	816	2 T17257	hypothetical prote
16	133.5	6.6	1634	1 JC5500	phosphoinositide 3
17	133.5	6.6	2472	2 A35715	fodrin alpha chain
18	133	6.5	408	2 T26940	hypothetical prote
19	131.5	6.5	303	2 S41754	CRKL protein - hum
20	131.5	6.5	2415	1 A33733	spectrin alpha cha
21	131.5	6.5	2477	1 S2CHA	spectrin alpha cha
22	130.5	6.4	1196	2 T41108	SH3-containing pro
23	129.5	6.4	303	2 S58352	SH2/SH3 adaptor pr
24	127.5	6.3	505	1 S24550	protein-tyrosine k
25	127.5	6.3	1249	2 A56511	myosin I myoA - Em
26	127	6.2	1244	2 S25327	cytoskeleton assem
27	125.5	6.2	517	2 A43807	protein-tyrosine k
28	125.5	6.2	1109	2 S53601	myosin-IC [similar
29	124	6.1	305	1 A49011	c-Crk - chicken

ALIGNMENTS

RESULT 1

A39249

neutrophil cytosol factor 1 - human

N:Alternate names: 47K autosomal chronic granulomatous disease protein; multicomponent
C:Species: Homo sapiens (man)

C>Date: 30-Jun-1993 #sequence-revision 30-Jun-1993 #text-change 18-Jun-1999

C:Accession: A35926; A39249; A54067; I59190; A32762; A41385

R:Rodaway, A.R.F.; Teahan, C.G.; Casimir, C.M.; Segal, A.W.; Bentley, D.L.

Mol. Cell. Biol. 10, 5388-5396, 1990

A:Title: Characterization of the 47-kilodalton autosomal chronic granulomatous diseases

A:Reference number: A35926; MUID:90377229

A:Accession: A35926

A:Molecule type: mRNA

A:Residues: 1-390 <ROD>

A:Cross-references: GB:M55067; GB:M38755; NID:gi189050; PIDN:AAA59901.1; PID:gl89051

R:Volpp, B.D.; Nauseef, W.M.; Donelson, J.E.; Moser, D.R.; Clark, R.A.

Proc. Natl. Acad. Sci. U.S.A. 86, 9563, 1989

A:Reference number: A39249

A:Contents: erratum

A:Accession: A39249

A:Molecule type: mRNA

A:Residues: 1-390 <VOI>

A:Cross-references: GB:M25665; GB:M26193; NID:gi189107; PIDN:AAA57209.1; PID:gl89108

R:Volpp, B.D.; Nauseef, W.M.; Donelson, J.E.; Moser, D.R.; Clark, R.A.

Proc. Natl. Acad. Sci. U.S.A. 86, 7195-7199, 1989

A:Title: Cloning of the cDNA and functional expression of the 47-kilodalton cytosolic

A:Reference number: A41385; MUID:89386707

A:Contents: annotation

A:Note: the sequence reported has been extensively revised in reference A39249

R:Lomax, K.J.; Leto, T.L.; Nunoi, H.; Gallin, J.I.; Malech, H.L.

Science 245, 409-412, 1989

A:Title: Recombinant 47-kilodalton cytosol factor restores NADPH oxidase in chronic g

A:Reference number: A32762; MUID:89332501

A:Contents: annotation

A:Note: the sequence reported has been extensively revised and now agrees with that s

R:Finan, P.; Shimizu, Y.; Gout, I.; Hsuan, J.; Truong, O.; Bucher, C.; Bennett, P.;

J. Biol. Chem. 269, 13752-13755, 1994

A:Title: An SH3 domain and proline-rich sequence mediate an interaction between two c

A:Reference number: A54067; MUID:94245680

A:Accession: A54067

A:Molecule type: protein

A:Residues: 8-16;44-52;71-77 <FIN>

A:Experimental source: differentiated HL-60 cells

R:Casimir, C.M.; Bu-Gharios, H.N.; Rodaway, A.R.F.; Bentley, D.L.; Rowe, P.; Segal, A.

Proc. Natl. Acad. Sci. U.S.A. 88, 2753-2757, 1991

A:Title: Autosomal recessive chronic granulomatous disease caused by deletion at a di

A:Reference number: I59190; MUID:91187670

A:Accession: I59190

A>Status: translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 14-24 <CAS>

A:Cross-references: GB:M50941; NID:gi189948; PIDN:AAA60086.1; PID:gl89949

parent can of 10/76734/

A:Note: a GT dinucleotide deletion at a GT-GT repeat causes a frameshift after residue 2
C:Comment: This protein is required for activation of the latent NADPH oxidase, which is
ous disease.

C:Genetics:

A:Gene: GDB:NCRL

A:Cross-references: GDB:120222; OMIM:233700

A:Map position: 7q11.23-7q11.23

A:Introns: 24/3

A:Note: the list of introns is incomplete

C:Superfamily: neutrophil cytosol factor 1; SH3 homology

C:Keywords: cytosol; neutrophil

F:163-210/domain: SH3 homology <SH31>

F:233-280/domain: SH3 homology <SH32>

Query Match 99.2%; Score 2018; DB 1; Length 390;

Best Local Similarity 98.7%; Pred. No. 1.1e-145;

Matches 385; Conservative 1; Mismatches 0; Indels 4; Gaps 1;

Qy 1 MGDTFIRHIALGFEKRFVPSQHYVYMFVWKQDLSEKVVYRRFTEIYEFHKTLEMEPI 60

Db 1 MGDTFIRHIALGFEKRFVPSQHYVYMFVWKQDLSEKVVYRRFTEIYEFHKTLEMEPI 60

Qy 61 EAGAINPENRIIPHPAPKWFQDQRAAENROGTLTEYESTLMSLPTKISRCPHLLDFFKV 120

Db 61 EAGAINPENRIIPHPAPKWFQDQRAAENROGTLTEYESTLMSLPTKISRCPHLLDFFKV 120

Qy 121 RPDDLKLPDQTKKPEYLMKPKDGKSTADITGPIILQTYRAIANYEKTSGSEMASTG 180

Db 121 RPDDLKLPDQTKKPEYLMKPKDGKSTADITGPIILQTYRAIANYEKTSGSEMASTG 180

Qy 181 DVVEVKESEGWFWCOMKAKRGWIPASFLPLDSPDETPEPNYAGEPYVAIKAYTAV 240

Db 181 DVVEVKESEGWFWCOMKAKRGWIPASFLPLDSPDETPEPNYAGEPYVAIKAYTAV 240

Qy 241 EGDEVSLLEGEAVEVIHKLDCG----KDDVTGYPSPMYLQSGODVSOAQRQIKRGAPP 296

Db 241 EGDEVSLLEGEAVEVIHKLDCG----KDDVTGYPSPMYLQSGODVSOAQRQIKRGAPP 296

Qy 297 RRSIRNAHSIHQSRKRLSDAYRRNSVRFLOQRRRQARPGQSPGSPLEEROTQRSK 356

Db 301 RRSIRNAHSIHQSRKRLSDAYRRNSVRFLOQRRRQARPGQSPGSPLEEROTQRSK 360

Qy 357 POPAVPPRPSADLILNRCSESTKRKLASAV 386

Db 361 POPAVPPRPSADLILNRCSESTKRKLASAV 390

RESULT 2

154525

C:Species: Mus musculus (house mouse)

C:Date: 02-Aug-1996 #sequence_revision 02-Aug-1996 #text_change 16-Jul-1999

C:Accession: 154525

R:Jackson, S.H.; Malech, H.L.; Kozak, C.A.; Lomax, K.J.; Gallin, J.I.; Holland, S.M.

Immunogenetics 39, 272-275, 1994

A:Title: Cloning and functional expression of the mouse homologue of p47phox.

A:Reference number: 154525; MUID:94164697

A:Accession: 154525

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-388 <RES>

A:Cross-references: GB:L11455; NID:g309422; PIDN:AAA50469.1; PID:g557868

C:Superfamily: neutrophil cytosol factor 1; SH3 homology

F:161-208/domain: SH3 homology <SH31>

Query Match 81.9%; Score 1666; DB 2; Length 388;

Best Local Similarity 80.2%; Pred. No. 5.7e-119;

Matches 312; Conservative 35; Mismatches 36; Indels 6; Gaps 3;

Qy 3 DTFIRHIALGFEKRFVPSQHYVYMFVWKQDLSEKVVYRRFTEIYEFHKTLEMEPIEA 62

Db 3 DTFIRHIALGFEKRFVPSQHYVYMFVWKQDLSEKVVYRRFTEIYEFHKTLEMEPIEA 62

Db 1 DTFIRHIALGFEKRFVPSQHYVYMFVWKQDLSEKVVYRRFTEIYEFHKTLEMEPIEA 60

Qy 63 GAINPENRIIPHPAPKWFQDQRAAENROGTLTEYESTLMSLPTKISRCPHLLDFFKV 122

Db 61 GEIHTENKRVIPHPAPKWFQDQRAAENROGTLTEYESTLMSLPTKISRCPHLLDFFKV 120

Qy 123 DDKLKLPDQTKKPEYLMKPKDGKSTADITGPIILQTYRAIANYEKTSGSEMASTG 182

Db 121 DDKLKLPDQTKKPEYLMKPKDGKSTADITGPIILQTYRAIANYEKTSGSEMASTG 180

Qy 183 VEVVEKSEGWFWCOMKAKRGWIPASFLPLDSPDETPEPNYAGEPYVAIKAYTAV 242

Db 181 VEVVEKSEGWFWCOMKAKRGWIPASFLPLDSPDETPEPNYAGEPYVAIKAYTAV 240

Qy 243 DEVSLLEGEAVEVIHKLDCG----KDDVTGYPSPMYLQSGODVSOAQRQIKRGAPP 297

Db 241 DMSLSEGEAVEVIHKLDCG----KDDVTGYPSPMYLQSGODVSOAQRQIKRGAPP 300

Qy 298 RRSIRNAHSIHQSRKRLSDAYRRNSVRFLOQRRRQARPGQSPGSPLEEROTQRSK 357

Db 301 RSTIRNAHSIHQSRKRLSDAYRRNSVRFLOQRRRQARPGQSPGSPLEEROTQRSK 359

Qy 358 QPAPVPPRPSADLILNRCSESTKRKLASAV 386

Db 360 QPAPVPPRPSADLILNRCSESTKRKLASAV 388

RESULT 3

T00056

hypothetical protein KIAA0418 - human

C:Species: Homo sapiens (man)

C:Date: 22-Jan-1999 #sequence_revision 22-Jan-1999 #text_change 05-Nov-1999

C:Accession: T00056

R:Ishikawa, K.; Nagase, T.; Nakajima, D.; Seki, N.; Ohira, M.; Miyajima, N.; Tanaka,

submitted to the EMBL Data Library, October 1997

A:Description: Prediction of the coding sequences of unidentified human genes. VIII.

A:Reference number: 214080

A:Accession: T00056

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-940 <ISH>

A:Cross-references: EMBL:AB007878; NID:d1179754; PIDN:BAA24848.1; PID:d1025770

A:Experimental source: brain

C:Genetics:

A:Note: KIAA0418

Query Match 13.5%; Score 275; DB 2; Length 940;

Best Local Similarity 26.9%; Pred. No. 8e-13;

Matches 70; Conservative 46; Mismatches 86; Indels 58; Gaps 7;

Qy 156 IILQTYRAIANYEKTSGSEMASTGDDVVVEKSPSGWFWCOMKAKRGWIPASFLPLD 215

Db 1 MLEQYVVSNTYKQENSELQAGEVDVIEKNESGWFWSTSEEGWVPATYLEAQNG 60

Qy 216 PDTEDEPPPNVAG--EPXVAIKAYTAVGEDEVSLLEGEAVEVIHKLDCG-----KDD 266

Db 61 TRDDSDINTSKTGEEKYVTVQPYTSQKDEIGFEKGVTVIEIRKNLEGGWVIRYLGR- 119

Qy 267 VTGYPPSPMYLQSGODVSOAQRQI-----KRGAPP----- 297

Db 120 --GWAPASYLKKAKDDLPTKRKNLAGPVEIIGNMEISNLNKKASGDKETPPABEGEGE 177

Qy 298 -----RSSIRNAHSIHQSRKRLSDAYRRNSVRFLOQRRRQARPGQSPG 343

Db 178 APIAKKETSLPILCNASNGSAGVDPDRTVSRLAQGS--PAVARIAPQRAQISSPNLTRP 235

Qy 344 SPLEEROTQRSKQPQAVPP 363

Db 236 PP-RRESSLGLQPKPPPEP 254

RESULT 4

T09194
adaptor protein intersectin - African clawed frog
C:Species: Xenopus laevis (African clawed frog)
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000
C:Accession: T09194
R:Yamabhai, M.; Hoffman, N.G.; Hardison, N.L.; McPherson, P.S.; Castagnoli, L.; Cesareni
J. Biol. Chem. 273, 31401-31407, 1998
A:Title: Intersectin, a novel adaptor protein with two eps15 homology and five src homol
A:Reference number: Z16605; MUID:99030416
A:Accession: T09194
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1270 <YAM>
A:Cross-references: EMBL:AF032118; NID:g2642624; PIDN:AAC73068.1; PID:g2642625
A:Experimental source: cell type oocyte
C:Function:
A:Description: involved in endocytosis
C:Keywords: endocytosis

Query Match 8.8%; Score 179.5; DB 2; Length 1270;
Best Local Similarity 21.1%; Pred. No. 2.2e-05;
Matches 93; Conservative 66; Mismatches 135; Indels 147; Gaps 23;

Qy 38 KVVYRRTEIYEFHKTLEKEMPIEAGINPENRIIPHPAPKWFQD----- 83
Db 733 KVVYR--ALYPFDARSHDEITIEPGDIIMVDE--SQTGEPCGLGGLKGTGWFPANYA 788
Qy 84 ORAENROGTLTEVCSTLMSLPT-KISRCP-----HLLDFKVRPDDLKLP 128
Db 789 ERMPSEFPSTTKPAAEETAKPTVHVAPSPVAPAAFTNTSTNSNNWADFSTWP----- 842
Qy 129 TDNOKKPE-----TYLMPKDGKS-----TATDITG---PIILQ-----T 160
Db 843 -TNNTDKVESDNWTWAAQPSLTVPASGQHQRAFTPATVTGSSPVLQCGKEVGLQ 901
Qy 161 YRAIANEKTSGSEMASTGSDVVEVEKESGWFQCMKARKGWPAPASFLEPLDSP----- 216
Db 902 AQALYPWRAKKDNHFNKNDVITVLEQDM-WWFEVQGGKGFPSYVVLISGLPKRS 960
Qy 217 --DET--EDP-----EPNYGEPVATKATAVGDEVSLLGEGAVEVIHKL 260
Db 961 TSIDSTSESASLKRVSPPAFKPAIOGEEYISMYTYESNQGLTFOQGLIIVVIRKDG 1020
Qy 261 DGWKDDV---GYPPSMYLO-----KSGQ-----DVSO-----AORIKRG 293
Db 1021 DWVTGTVEKTVGPSNIVYRPKDSKSEAGSGKTSGLKKPEIAQVIASVYATAPQLTL- 1079
Qy 294 APPRRSSIRNAH-----SIHRSRKRSLSDAYRRNSVRFLOQRRRQARPQSPGSL 346
Db 1080 APGQLILIRKNKPGGWGEGLQARGKKR-QIGWFPANYVKLL-----SPGT-- 1124
Qy 347 EERQRTQSRKQPAVPPRPSA 367
Db 1125 -----NKSTPEPKPTS 1137

RESULT 5
S39768
neutrophil cytosol factor p40 - human
C:Species: Homo sapiens (man)
C:Date: 19-May-1994 #sequence_revision 17-Nov-1995 #text_change 07-May-1999
C:Accession: S39768; S39769
R:Wientjes, F.B.; Hsuan, J.J.; Totty, N.F.; Segal, A.W.
Biochem. J. 296, 557-561, 1993
A:Title: p40(phox), a third cytosolic component of the activation complex of the NADPH o
A:Reference number: S39768; MUID:94107216
A:Accession: S39768
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-339 <WIE>
A:Accession: S39769
A:Molecule type: protein

A:Residues: 52-92;115-125;131-145;146,176-186;192-213;228-234;241-263;297-308;317-322
C:Superfamily: SH3 homology
C:Keywords: cytosol
F:177-224/Domain: SH3 homology <SH3>

Query Match 8.3%; Score 169.5; DB 2; Length 339;
Best Local Similarity 24.4%; Pred. No. 2e-05;
Matches 76; Conservative 48; Mismatches 116; Indels 71; Gaps 12;

Qy 15 EKRFPVPSOHVYVMEFLVKWQDLSEKVVYRRFTETIYEFHKTLEKEMPIEAGINPENR---- 70
Db 31 EKRGETS-HFVFVTEVTKGSKYLIYRRYRQFHALQSKLEERF-----GPDSKSSAL 82
Qy 71 --IIPHPAPKWFQD--ORAAENROGTLTEVCSTLMSLPTKISRCPHLLDFFKVRP-DDL 125
Db 83 ACTLPTLPA-KVYGVKQEIEMRIPALNAYMKSLSLPVMVLMDEDVRIFFYQSPYDSE 141
Qy 126 KLPTDNQTKPKETYLMPKDGKSTATDITGPIILQYIRAIANYEKTSGSEMALS--TGDVV 183
Db 142 QVPAIRRLRPRT----RKVKSVPQGSVDMAAPRAEALFDFTGSKLELNFKAGDVI 197
Qy 184 EVVEKESGWFQCMKARKGWPAPASFLEPLDSPDETP----- 222
Db 198 FLLSRINKDWLEGTVRGATGIRPLSFVKILKDPPEEDPTNMLRCYVYEDTISTIKDIAV 257
Qy 223 EPNVAGEPY-----VAIKAYTAVGDEVSLLGEGAVEVI-----H 257
Db 258 EEDLSSTPLKDLLELTRREFQREDIALN-YDAEGDLVRLLSDEDVALMVRQARGLPSQ 316
Qy 258 KLLDGWKDDVT 268
Db 317 KRLFPWKLIHT 327

RESULT 6
T13151
adaptor protein CMS - human
C:Species: Homo sapiens (man)
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 17-Nov-2000
C:Accession: T13151; T08754
R:Kirsch, K.H.; Georgescu, M.M.; Ishimaru, S.; Hanafusa, H.
Proc. Natl. Acad. Sci. U.S.A. 96, 6211-6216, 1999
A:Title: CMS: An adaptor molecule involved in cytoskeletal rearrangements.
A:Reference number: Z17608; MUID:99272673
A:Accession: T13151
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-639 <KIR>
A:Cross-references: EMBL:AF146277; NID:g4960046; PID:g4960047; PIDN:AAD34595.1
R:Wambutt, R.; Heubner, D.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, March 1999
A:Reference number: Z16471
A:Accession: T08754
A:Molecule type: mRNA
A:Residues: 548-639 <WAM>
A:Cross-references: EMBL:AL050105
A:Experimental source: adult uterus; clone DKFZp586H0519
C:Genetics:
A:Note: DKFZp586H0519.1
C:Complex: homodimer
C:Function:
A:Description: probably functions as a scaffolding molecule with a specialized role 1
C:Keywords: coiled coil; homodimer

Query Match 8.1%; Score 164.5; DB 2; Length 639;
Best Local Similarity 22.9%; Pred. No. 0.00011;
Matches 86; Conservative 50; Mismatches 127; Indels 113; Gaps 17;

Qy 45 TEIYEFHKTLEKEMPIEAGINPENRIIPHPAPKWFQDQRAAENRQGTPLTETCYSTLMSL 104
Db 168 TDDGETHEAQDDSETVLAGPTSP----IPSL-----GNVSETASGSVTQ 207

Qy	105	PTKISRCHLLDFFKVRPDDLKLPT---	DNQTKKPYTLM-----PK--	DGKSTATDITG	154
		: :	: :	:	:
Db	208	PKKI-RGIGFGDIFKEGSVLRTRTSSETEEEKPEKPLILQSLGPKTQSVEITKTDTGS	266		
Qy	155	PIILOTY-RAIANYEKTSSEMALSTGDUVVVEVK--	SESGWFCOMKAKRGWIPASFLE	211	
		: :	: :	:	:
Db	267	KIRAKEYCRTLFAYEGTNEDELFTFKEGEIIHLISKETGEAGWRGELNGKEGVFPDNFAV	326		
Qy	212	PL-----DSPDETEDPENYAGPB---YVAIKAYTAV--	ECDEVISLLEGEAVEVIHL	259	
		: :	: :	:	:
Db	327	QINELDKFPRKPPPPAPAKAPKPELIAAEKKYSLSAPEKDEKSTLE-	-----	376	
Qy	260	LDGWKDWDVTGYFYSMWLQKSGODVSQAQRCIKGAPRASS--	IRNAHSITHORSKRRLSQ	317	
		: :	: :	:	:
Db	377	-----QKPSKFAA-PQVPPPKPPTPTKASNLRRSSGVYP-	-----	410	
Qy	318	DAYRNRSRVFLQORRRQARPGPOSP-	-----GSPLEEEROTQRSKPQPAPPPRP	365	
		: :	: :	:	:
Db	411	-----KRPEKVPVPPPPIAKINGEVBSSISKFEETEPYSKLKLDSQLPLRP	456		
Qy	366	SA----DLILNRCEST	378		
		: :	: :	:	:
Db	457	KSVDFDSLTVRTSKET	472		

```

RESULT      7
TI13055
dynamlin associated protein isoform 160-2 - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C:Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000
C:Accession: Tl13055
R:Roos, J.; Kelly, R.B.
J. Biol. Chem. 273, 19108-19119, 1998
A:Title: Dap160, a neural-specific Eps15 homology and multiple SH3 domain-containing protein
A:Reference number: Z17594; MUID:98354647
A:Accession: Tl13055
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-1011 <R00>
A:Cross-references: EMBL:AF054612; NID:g2996029; PID:g2996030; PIDN:AAC39139.1
C:Genetics:
A:Gene: Dap160
A:Cross-references: FlyBase:FBgn0023388

```

```
Query Match      8.1%; Score 164; DB 2; Length 1011;
Best Local Similarity 22.4%; Pred. No. 0.00024;
Matches 79; Conservative 42; Mismatches 121; Indels 110; Gaps 14;
```

```

Qy 104 LPTKISRCPLHLLFFKVRPD---DLKLTDNQTTPETLMPDKGSKSTATDITGPILQ- 159
|   | : : |   | : : |   | : : |   | : : |   | : : |   | : : |
Db 594 LSLALITKCEDLYKEYDVQRTSVLELAYNRKNETSVSAM---DTGSSSAWEETGTTVDP 650
Qy 160 -----TTRATANYEKTSGSEMAIUSTGDV---VEVVEKSE 190
|   | : : |   | : : |   | : : |   | : : |   | : : |   | : : |
Db 651 YAVASNDISALAAPAYDLGPAPEGVFKYQAVYEFNARNAEITTFPGDPIILLPLEQNAE 710
Qy 191 SGWWFCOMKAKRWIPASTLEPLDSPD-----ETEDPPEPVAGEP---- 230
||   || : : || : : || : : || : : || : : || : : || : : ||
Db 711 PGWLAGEINGHTCMFPFESYVELEVGEVAPAAVEAPDAQVAODYNDINTSSTPAASA 770
Qy 231 -----YVAIKAYTAVEGDEVSLLEGEAVEVTHKLLDGWKDDV---TCYFFSMYLQ 277
|   | : : |   | : : |   | : : |   | : : |   | : : |   | : : |
Db 771 DLTAAGDVEYYIAAPYESAEBGLSFSGAEWVMJLKKEGWNTGTIGRTGMFFSNVYQ 830
Qy 278 KSGQDVYSQAQ-----RQIKRGAPPRRSRIR-----NAHSIHORS-----KKR- 314
|   | : : |   | : : |   | : : |   | : : |   | : : |   | : : |
Db 831 KA--DVGCTASTAAEPVESLDQGMRAKRSEIAQVIAPYEATSTEQLSTRGOLIMRKKT 888
Qy 315 -----LSQDAYRR-----NSVRFTLOQRRRQARPQPSPGPSLEPERQ 351
Db 889 DSGWWECELAQARRRQIOGFWPATYVKVLGGRRNSGRNTPVSGRIENITEO 939

```

RESULT
T13053
8

dynamid associated protein isoform Dap160-1 - fruit fly (*Drosophila melanogaster*)
C/Species: *Drosophila melanogaster*
C/Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000
C/Accession: T13053
R;Roos, J.; Kelly, R.B.
J. Biol. Chem. 273, 19108-19119, 1998
A/Title: Dap160, a neural-specific Eps15 homology and multiple SH3 domain-containing
A/Reference number: Z17594; MUID:98334647
A/Accession: T13053
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: mRNA
A/Residues: 1-1094 <R00>
A/Cross-references: EMBL:AF053957; NID:g2984714; PID:g2984715; PIDN:AAC39138.1
C/Genetics:
A/Gene: Dap160
A/Cross-references: FlyBase:FBgn0023388

Query Match 7.9%; Score 161; DB 2; Length 1094;
Best Local Similarity 21.3%; Pred. No. 0.00045;
Matches 77; Conservative 50; Mismatches 125; Indels 110; Gaps 15;

Qy	104	LPTKISRCPHLLDFKVRPD---	DLKLPTDQTKKKPETYLMPKDGKSTATDITGPILQ-	159
		: : : : : :	: : : : : :	
Db	594	LSALITKCEDLYKEYDVORTSVLELKYNRRNKTSSVSAW---	DTGSASAMEETGTVPDP	650
Qy	160	-----TYRAIANYEKTSCEMALSTGDV--VEVVEKSE	190	
		: : : : : :	: : : : : :	
Db	651	YAVANSDISALAAPVDLGGPAPEGVKYQAUVFENARNAEETFPVGDIILVPLEQNAE	710	
Qy	191	SGMWFCOMKAKRGWIPASFLEPLDSPD-----ETEDPEPNYAGEP----	230	
Db	711	PGWLAGEINGHTGWFPESYVEKLEVGEVAPVAAVEAPVDAQVADTYNDNINTSSIIPAASA	770	
Qy	231	-----YVAIKATYAVEGDEVSLLEGEAVEVTHKLLDGNKDDV---TGYPFSMYLQ	277	
		: : : : : :	: : : : : :	
Db	771	DLTAAGDVEYYIAAPYESAEGLDFSAGEMVMVIKKEGEWMTGTIGRTGMFPFNYSVQ	830	
Qy	278	KSGODVSQAORQIKRGAPPRRS---	SIRNAHSITHORSRRLSQDAYRRNSVFELQORRR	333
		: : : : : :	: : : : : :	
Db	831	KA--DVGTAS---TAAAEPVESLDQETTLNGNAAYTAAPVEAQEQVYQLPVO-----	878	
Qy	334	QARPQPQ---SPGSPLLEE-----ERQTRSXPQAVPPRPSPSADLIILNCRSE	376	
		: : : : : :	: : : : : :	
Db	879	--EPSEQPISSPGVGAEEAHELDLTVEVSQINTOSKTQSSEPASESY-SRP-----MSRTSS	930	
Qy	377	ST	378	
Db	931	MT	932	

RESULT 9

JC7191

85K c-Cbl-interacting protein, CIN85 - human

C:Species: Homo sapiens (man)

C:Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 23-Mar-2001

C:Accession: JC7191

R:Takae, H.; Watanabe, S.; Takeda, K.; Yu, Z.X.; Iwata, N.; Kajigaya, S. Biochem. Biophys. Res. Commun. 268, 321-328, 2000

A:Title: Cloning and characterization of a novel adaptor protein, CIN85, that interacts with c-Cbl

A:Reference number: JC7191; MUID:20145431; PMID:10679202

A:Accession: JC7191

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-665 <YAK>

A:Cross-references: GB:AF230904; NID:g7188748; PID:g7188749

C:Genetics:

A:Gene: CIN85

Query Match 7.4%; Score 150.5; DB 2; Length 665;
Best Local Similarity 24.3%; Pred. No. 0.0014;
Matches 43; Conservative 34; Mismatches 51; Indels 49; Gaps 7;
QY 166 NYKTSSEALSTGDDVVVEVVEKSGWFCQMKAKRGWIPASLEPLDSDPDEDEPN 225
DB 9 DYQAQHDDELTSVGEITIRKEDGGWEGQINGRRGLFPDNFVREI-RKEMKKDLTN 67
QY 226 YAGEP-----YVAIKAYTAVEGDEVSLLEGEAVE 254
DB 68 KAPEKPLHEVPNGSNLSSETILTNKRGERRRRRCQVAF-SYLPQNDDELEKVGDIIE 126
QY 255 VIHLLDGWQDDV-----TGVPFPMYLOK-SGQD-----VSQAQRQIKRGAPPRRSIR 302
DB 127 VVGEVGEWGVNGKGTGMPFNSFIKELSGESDELGISQDEQLSK-----SSLR 176
RESULT 10
T42526
hypothetical protein - fission yeast (Schizosaccharomyces pombe) (fragment)
C:Species: Schizosaccharomyces pombe
C:Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 21-Jul-2000
C:Accession: T42526
R:Yoshioaka, S.; Kato, K.; Nakai, K.; Okayama, H.; Nojima, H.
DNA Res. 4, 363-369, 1997
A:Title: Identification of open reading frames in Schizosaccharomyces pombe cDNAs.
A:Reference number: 217323; MUID:98162722
A:Accession: T42526
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-290 <YOS>
A:Cross-references: EMBL:D89164; NID:g1749535; PIDN:BAA13826.1; PID:g1749536
A:Experimental source: strain PR745
Query Match 7.2%; Score 147.5; DB 2; Length 290;
Best Local Similarity 20.7%; Pred. No. 0.00074;
Matches 59; Conservative 56; Mismatches 113; Indels 57; Gaps 11;
QY 5 FIRHALLGFEKRVPSQHYVYVFLVWQDLSEKVVVRRFTEIYEF-HKTKEMFPPEAG 63
DB 51 FLRGLLFLYQKPRSPRSHLLHLLFLQFLNLHLLLLYLLKLLQFLNHLPLLYLRLFLQ 110
QY 64 AINPRTIPLHAPKWFQDQRAAENRQGLTEYCSLTMSLPTKISRCPHLLDFEKVRPD 123
DB 111 FLKPP--AVPVVPEP-----AGQLNE-----PV 131
QY 124 DLKLPDNTQKKPETYLMPKDGKSTATDITGPIILOTYRAIANYEKTSGSEMASTGDDV 183
DB 132 VPPLPPHDETQEPQV-----GGDVKATEHTQPTKTPAI-VIYDYSPEENEIELVENEQI 185
QY 184 EVVEKSGWFWFCQ-MKAKRGWIPASLEPLDSDPDEDE--PEPNYAGEPVVAIKA---Y 237
DB 186 QILEFVDGWMWLGNSGQGOQLFPNSYVE-ITGPNETANNPPAPQAGPGKSKAIYDI 244
QY 238 TAVEGDEVSLLEGEAVEVIHLLDGWQDDV-----TGVPFPMYLOK 278
DB 245 QAQEDNELSFDEDELIANVDCVDPNWEGECHGRGLFPNSYVEE 289
RESULT 11
T50995
related to cytoskeleton assembly control protein SLAL [imported] - Neurospora crassa
N:Alternate names: protein B7F18.140
C:Species: Neurospora crassa
C:Date: 21-Jul-2000 #sequence_revision 21-Jul-2000 #text_change 21-Jul-2000
C:Accession: T50995
R:Schultze, U.; Aign, V.; Hoheisel, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatura,
submitted to the Protein Sequence Database, July 2000
A:Reference number: T25286
A:Accession: T50995
A:Status: preliminary

A:Molecule type: DNA
A:Residues: 1-1119 <SCH>
A:Cross-references: EMBL:AL389891; GSPDB:GN00116; NCSP:B7F18.140
A:Experimental source: BAC clone B7F18; strain OR74A
C:Genetics:
A:Gene: NCSP:B7F18.140
A:Map position: 6
A:Introns: 66/3; 123/2; 495/1
Query Match 7.1%; Score 145; DB 2; Length 1119;
Best Local Similarity 25.6%; Pred. No. 0.0076;
Matches 62; Conservative 31; Mismatches 95; Indels 54; Gaps 11;
QY 158 LOTYRAIANYEKTSGSEMASTGDDVVVEVVEKSGWFCQMKAKR-----GWIPASF 209
DB 4 LGVYRAIYDTPOGEGELTISEGDILYVLEKSDQEDDWKAKKANAAADDEPVGILPNNY 63
QY 210 LEPLDSDPDEDEPNYAGEPVVAIKAYTAVEGDEVSLLEGEAVEVIHKLDDGM-----KD 265
DB 64 IE-----EAKPVSVAR-----ALYEYTRQTDLELSFPEDAQLSVFOTSDPWLVLVGH 111
QY 266 DVTGYFPMYLOKSGQD-----VSQAQRQIKRGAPPRRSIRNAH--SIHQSRKRLSQ 317
DB 112 GDYGFAPANYTESVPSEPTPVAASNPAATAAGVMAHRSSFPQPPAPISVPQPOQSYAS 171
QY 318 DAYRRNSVRFLLQRRRRQARPQPSP--CSPLEEEROTQRSKPOPA-----VPPR 364
DB 172 EDY-ENEVR-----SPPLPSRPRGDSQIAPEQKSVRPVPPPAQSAHAHTDDDGISPR 222
QY 365 PS 366
DB 223 TS 224
RESULT 12
T31504
hypothetical protein Y116A8C.36 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C:Accession: T31504
R:McMurray, A.
submitted to the EMBL Data Library, October 1999
A:Reference number: Z21041
A:Accession: T31504
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1097 <WIL>
A:Cross-references: EMBL:AL117204; PIDN:CAB55138.1; CESP:Y116A8C.36
A:Experimental source: clone Y116A8C
C:Genetics:
A:Gene: CESP:Y116A8C.36
A:Introns: 50/3; 245/3; 411/3; 486/3; 697/1; 820/3; 926/2
Query Match 7.0%; Score 142.5; DB 2; Length 1097;
Best Local Similarity 26.3%; Pred. No. 0.011;
Matches 44; Conservative 33; Mismatches 67; Indels 23; Gaps 7;
QY 140 LMPKDGKSTATDITGPIILOTYRAIANYEKTSGSEMASTGDDVVVEVVEKSGWFCQMK 199
DB 880 VVPSDVTLQASE-TAP-QQQLYTVIYDFEAVETTDLALHVGDTILVLEKNDP-WNKGRCN 936
QY 200 AKRGWIPASFLEPLDSDPDEDEP-----EPNYAGEPVVAIKA-----YTAVEGDEVSLLE 249
DB 937 GREGIFPANYVE--TSVQOAGDPPTPTQAPTAPAPTTLCEAKVVVDFVASAPNOLGIKV 994
QY 250 GEAVEVIHKLDDGWDK-----DVTGVFPMYLOKSGQDVSQAQR 288
DB 995 GEIVKIREKSAAGWWEGLIRNGKPIAGWFFGGEYVKVLEEAASPATR 1041
RESULT 13

[illegible]

Search completed: January 11, 2002, 09:08:16
Job time: 142 sec

Query Match	6.7%	Score 136;	DB 4;	Length 50;
Best Local Similarity	92.6%;	Pred. No. 0.00048;		
Matches 25;	Conservative	2;	Mismatches 0;	Gaps 0;
			Indels	0;

QY 1 MGD F I R H I A L L G F E K R V P S Q H Y V Y M 27
| | | | | | | | | | | | | | | | | | | | | | : :

THIS PAGE BLANK (USPTO)

Printed 10/767341

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 11, 2002, 09:05:54 ; Search time 25.33 seconds
(without alignments)
1128.792 Million cell updates/sec

Title: US-09-820-005-2
Perfect score: 2035
Sequence: 1 MGDFTIRHALLGFEKRFVP.....ADLIILRCSESTKRKLASAV 386

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :			
A_Geneseq_1101.*			
1:	/SID88/gcgdata/geneseq/geneseqp/AA1980.DAT.*		
2:	/SID88/gcgdata/geneseq/geneseqp/AA1981.DAT.*		
3:	/SID88/gcgdata/geneseq/geneseqp/AA1982.DAT.*		
4:	/SID88/gcgdata/geneseq/geneseqp/AA1983.DAT.*		
5:	/SID88/gcgdata/geneseq/geneseqp/AA1984.DAT.*		
6:	/SID88/gcgdata/geneseq/geneseqp/AA1985.DAT.*		
7:	/SID88/gcgdata/geneseq/geneseqp/AA1986.DAT.*		
8:	/SID88/gcgdata/geneseq/geneseqp/AA1987.DAT.*		
9:	/SID88/gcgdata/geneseq/geneseqp/AA1988.DAT.*		
10:	/SID88/gcgdata/geneseq/geneseqp/AA1989.DAT.*		
11:	/SID88/gcgdata/geneseq/geneseqp/AA1990.DAT.*		
12:	/SID88/gcgdata/geneseq/geneseqp/AA1991.DAT.*		
13:	/SID88/gcgdata/geneseq/geneseqp/AA1992.DAT.*		
14:	/SID88/gcgdata/geneseq/geneseqp/AA1993.DAT.*		
15:	/SID88/gcgdata/geneseq/geneseqp/AA1994.DAT.*		
16:	/SID88/gcgdata/geneseq/geneseqp/AA1995.DAT.*		
17:	/SID88/gcgdata/geneseq/geneseqp/AA1996.DAT.*		
18:	/SID88/gcgdata/geneseq/geneseqp/AA1997.DAT.*		
19:	/SID88/gcgdata/geneseq/geneseqp/AA1998.DAT.*		
20:	/SID88/gcgdata/geneseq/geneseqp/AA1999.DAT.*		
21:	/SID88/gcgdata/geneseq/geneseqp/AA2000.DAT.*		
22:	/SID88/gcgdata/geneseq/geneseqp/AA2001.DAT.*		

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	2006	98.6	390	16	AA83825
2	691	34.0	128	22	AAG64031
3	688	33.8	128	22	AAG64033
4	593	29.1	141	21	AAB53681
5	305	15.0	64	22	AAG64032
6	295	14.5	64	22	AAG64034
7	261	12.8	968	22	AAU14174
8	188	9.2	509	17	AAW05399
9	187	9.2	438	21	AAB40983
10	183	9.0	1683	21	AAV71160
11	177.5	8.7	1197	21	AAV57445

12	177.5	8.7	1658	21	AAV57450	Mouse Ese2L protei
13	176.5	8.7	248	17	AAW05396	Human SH3P18 prote
14	172	8.5	1214	21	AAV57444	Mouse Esel protein
15	168.5	8.3	1715	21	AAV57449	Mouse EsellL protei
16	167	8.2	1215	20	AAV32156	Human SH3D1A prote
17	166	8.2	1220	20	AAV32155	Human SH3D1A prote
18	157	7.7	462	17	AAW05395	Human SH3P17 prote
19	157	7.7	641	20	AAV32158	Human SH3D1A prote
20	157	7.7	1144	20	AAV32154	Human SH3D1A prote
21	156.5	7.7	709	21	AAV19691	Human NNG2 (long f
22	150.5	7.4	665	21	AAV19690	Human NNG2 (short
23	150	7.4	665	21	AAV19685	Rat NNG2 (RUK1) pr
24	147.5	7.2	324	18	AAW26497	CD2-associated int
25	147.5	7.2	324	18	AAW25117	CD2-associated int
26	147.5	7.2	324	22	AAW80421	CD2-associated int
27	147.5	7.2	324	22	AAW66392	Human CD2 associat
28	144.5	7.1	635	19	AAW48898	Candida albicans C
29	144	7.1	464	18	AAW26496	CD2-associated int
30	144	7.1	464	18	AAW25116	CD2-associated int
31	144	7.1	464	20	AAW80420	CD2-associated int
32	144	7.1	464	22	AAW66391	Human CD2 associat
33	141	6.9	553	18	AAW26495	CD2-associated int
34	141	6.9	553	18	AAW25115	CD2-associated int
35	141	6.9	553	20	AAW80419	CD2-associated int
36	141	6.9	553	22	AAW66390	Human CD2 associat
37	141	6.9	659	19	AAW37724	CD2-associated int
38	137.5	6.8	397	20	AAV28292	Amino acid sequenc
39	137.5	6.8	416	20	AAV28291	Amino acid sequenc
40	137.5	6.8	416	21	AAV68782	Amino acid sequenc
41	134.5	6.6	1035	22	AAW43519	Human polyptide
42	133.5	6.6	370	21	AAV99428	Human PRL1431 (UNQ
43	133.5	6.6	370	22	AAW66177	Protein of the inv
44	133.5	6.6	1200	21	AAV19313	Amino acid sequenc
45	132.5	6.5	415	19	AAW71595	Murine tyrosine ph

ALIGNMENTS

RESULT	1
AA83825	
ID	AA83825 standard; Protein; 390 AA.
XX	
AC	AA83825;
XX	
DT	15-FEB-1996 (first entry)
XX	
DE	p47(phox) protein.
XX	
KW	p47(phox); p67(phox); NADPH oxidase complex; proline-rich region; PTPase;
KW	SH3 domain; inhibition; dynamin; cytochrome b245; reperfusion injury;
KW	septic shock; arthritis; asthma; vinculin; inflammatory bowel disease;
KW	adult respiratory distress syndrome; ischaemic heart disease.
XX	
OS	Homo sapiens.
XX	
EH	Key
FT	Misc-difference 22
FT	/label= OTHER
FT	/note= "designated O in specification"
FT	Misc-difference 197
FT	/label= OTHER
FT	/note= "designated O in specification"
FT	360..373
FT	/note= "proline-rich region used to design inhib"
XX	
PN	GB2285047-A.
XX	
PD	28-JUN-1995.
XX	
PF	07-DEC-1994; 94GB-0024674.
XX	
PR	07-JAN-1994; 94GB-0000248.

PR 21-DEC-1993; 93GB-0026083.
XX (LUDW-) LUDWIG INST CANCER RES.
PA (YAMA) YAMANOUCHI UK LTD.
XX
PI Finan PM, Gout IT, Kellie S, Shimizu Y, Waterfield MD;
XX WPI; 1995-217703/29.
DR
XX
XX New polypeptide(s) with antiinflammatory action - inhibit NADPH
PT oxidase system.
PT
XX
PS Example 1; Fig 1; 17pp; English.
XX
XX The protein p47(phox) (AAR83825) interacts with protein p67(phox) in the
CC NADPH oxidase complex via a C-terminal proline-rich region of p47(phox)
CC binding to at least one C-terminal SH3 domain of p67(phox). The
CC peptides AAR83814-24 were derived from the proline-rich region of
CC p47(phox) and show inhibitory activity towards the binding of p47(phox)
CC to p67(phox) as compared to the proline-rich region of a range of other
CC proteins (AAR83826-30). The inhibitory peptides can be used for the
CC treatment of chronic or acute inflammatory diseases e.g. septic shock,
CC arthritis, asthma, adult respiratory distress syndrome, ischaemic heart
CC disease, reperfusion injury or inflammatory bowel disease.
XX
SQ Sequence 390 AA;

Query Match 98.6%; Score 2006; DB 16; Length 390;
Best Local Similarity 98.2%; Pred. No. 8.5e-187;
Matches 383; Conservative 1; Mismatches 2; Indels 4; Gaps 1;
Qy 1 MGDTFIRHALLGKEKRPVPSOHVYVYMLVKWDLSEKVVYRRTEIYEFHKTLEMPPI 60
Db 1 mgdtfirhiallgfkrfvpsghvymflvkwdlsekvvrrrfteyefhktikempfi 60
Qy 61 EAGAINPENRIIPLPAPKWFQDGOAAENRQGTLETCSTLMSLPTKISRCPHLLDFFKV 120
Db 61 eagainpenriiiphlpapkwfdgqraaenrggtlteycstlmslptksrphlldffkv 120
Qy 121 RPDDLKLPDQTKPEYILMPKDGKSTATDITGPILQTYRAITANYEKTSGSEMA1STG 180
Db 121 rpddliklpdntqtkpetylmpkdgkstatdgtgpiilqtyraiaadyektsgsema1stg 180
Qy 181 DVEVEKSESGWVFCOMKARGWTPASFLPDLSPDTEPEPNYAGEPVVAIKAYTAV 240
Db 181 dvvevkesesgwvfcxmkaqrwpasflpdlspdtepepnayagepvvaikaytav 240
Qy 241 EGDEVSLLEGEBAVEVIHKLDDGW----KDDVTGYFPFMYLQKSGQDVSAQORQIKRGAPP 296
Db 241 egdevsllegeaveviahklldgwvvrkddvtgyfypsnylqksggdvsqagrqikrgapp 300
Qy 297 RRSSTIRNAHSTHQRKFLSODAYRNSVRLQORRQARPCQSPGSPLEERQTQRSK 356
Db 301 rrssirnahshqrsrksldqayrrnsvrlqrrrqarpgpdpqspgleeertqrsk 360
Qy 357 PQPAPVPPRPSADLILNRCSESTKRLASAV 386
Db 361 qpapvpprpsadlilnrcsestkrklasav 390

RESULT 2
AAG64031
ID AAG64031 standard; protein; 128 AA.
XX
AC AAG64031;
XX
DT 07-SEP-2001 (first entry)
XX
DE Human PX domain.
XX
KW Human; antiinflammatory; protein coordinate data; chemical shift data;
KW nuclear magnetic resonance; NMR; structural coordinate data;
XX

KW PX domain; reduced nicotinamide adenine dinucleotide phosphate; NADPH;
KW NADPH oxidase; superoxide production; inflammatory disease.
XX
OS Homo sapiens.
XX WO200142453-A1.
XX PN
XX PD 14-JUN-2001.
XX
XX 01-DEC-2000; 2000WO-JP08501.
XX PF
XX PR 06-DEC-1999; 99JP-0346193.
XX PA (BIOM-) BIOMOLECULAR ENG RES INST.
XX
XX Kohda D, Hiroaki H, Sumimoto H;
XX WPI; 2001-381679/40.
XX
XX Regulation of function of a protein containing a PX domain for
XX controlling an inflammatory response to disease or trauma, comprises
XX using nuclear magnetic resonance (NMR) chemical shift and structural
XX coordinate data -
XX
XX Claim 14; Page 182-183; 195pp; Japanese.
XX
XX The present sequence is provided in a specification relating to the use
XX of chemical shift data from nuclear magnetic resonance (NMR) and
XX structural coordinate data of the PX domain of a protein to search for,
XX evaluate, design and identify variant PX domain sequences for
XX controlling the function of proteins containing the PX domain. It also
XX relates to compounds promoting the binding of substances to the PX
XX domain, and compounds inhibiting the binding of substances to the PX
XX domain. The invention can be used in the identification of substances
XX that can be used in the regulation of reduced nicotinamide adenine
XX dinucleotide phosphate (NADPH) oxidase activity, which is involved in
XX superoxide production in inflammatory response in disease or trauma.
XX
SQ Sequence 128 AA;
Query Match 34.0%; Score 691; DB 22; Length 128;
Best Local Similarity 100.0%; Pred. No. 2e-59;
Matches 128; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MGDTFIRHALLGKEKRPVPSOHVYVYMLVKWDLSEKVVYRRTEIYEFHKTLEMPPI 60
Db 1 mgdtfirhiallgfkrfvpsghvymflvkwdlsekvvrrrfteyefhktikempfi 60
Qy 61 EAGAINPENRIIPLPAPKWFQDGOAAENRQGTLETCSTLMSLPTKISRCPHLLDFFKV 120
Db 61 eagainpenriiiphlpapkwfdgqraaenrggtlteycstlmslptksrphlldffkv 120
Qy 121 RPDDLKLP 128
Db 121 rpddliklp 128
RESULT 3
AAG64033
ID AAG64033 standard; protein; 128 AA.
XX
AC AAG64033;
XX
DT 07-SEP-2001 (first entry)
XX
DE Human PX domain-related protein.
XX
KW Human; antiinflammatory; protein coordinate data; chemical shift data;
KW nuclear magnetic resonance; NMR; structural coordinate data; PX domain;
KW reduced nicotinamide adenine dinucleotide phosphate; NADPH;
KW NADPH oxidase; superoxide production; inflammatory disease.
XX

```
OS Homo sapiens.
XX WO200142453-A1.
XX 14-JUN-2001.
XX
XX 01-DEC-2000; 2000WO-JP08501.
XX PF
XX 06-DEC-1999; 99JP-0346193.
XX PR
XX (BIOM-) BIOMOLECULAR ENG RES INST.
XX PA
XX Kohda D, Hiroaki H, Sumimoto H;
XX PI
XX WPI; 2001-381679/40.
XX DR
XX
XX Regulation of function of a protein containing a PX domain for
XX PT controlling an inflammatory response to disease or trauma, comprises
XX PT using nuclear magnetic resonance (NMR) chemical shift and structural
XX PT coordinate data -
XX PS
XX Example 3; Page 186; 195pp; Japanese.
XX
XX The present sequence is provided in a specification relating to the use
XX CC of chemical shift data from nuclear magnetic resonance (NMR) and
XX CC structural coordinate data of the PX domain of a protein to search for,
XX CC evaluate, design and identify variant PX domain sequences for
XX CC controlling the function of proteins containing the PX domain. It also
XX CC relates to compounds promoting the binding of substances to the PX
XX CC domain, and compounds inhibiting the binding of substances to the PX
XX CC domain. The invention can be used in the identification of substances
XX CC that can be used in the regulation of reduced nicotinamide adenine
XX CC dinucleotide phosphate (NADPH) oxidase activity, which is involved in
XX CC superoxide production in inflammatory response in disease or trauma.
XX
XX Sequence 128 AA;
SQ
Query Match 33.8%; Score 688; DB 22; Length 128;
Best Local Similarity 99.2%; Pred. No. 4e-59;
Matches 127; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Oy 1 MGDFTIRHIALGFEKRFVPSQHYVYMFVLVKWQDLSEKVVYRRFTEIYEFHKLKEMFPI 60
Db 1 mgdtfirhiallgfekrfvpsqhyvymflvkwdlsekvvyrfrteiyefhklkemfpi 60
Oy 61 EGAINPENRIIPHLPAKWFQDQRAAENRGTLTEYCSLTMSLPTKISRCPHLLDFKVP 120
Db 61 esgainpenriiphlpapkwfqdgqraaenkgtlteycstlmslptkiskrcphlldfkvp 120
Oy 121 RPDDLKLP 128
Db 121 rpddliklp 128
RESULT 4
AAB53681
ID AAB53681 standard; Protein; 141 AA.
XX
XX AAB53681;
XX
XX 09-MAR-2001 (first entry)
XX
XX Human colon cancer antigen protein sequence SEQ ID NO:1221.
XX
XX Human; colon cancer; colon cancer antigen; diagnosis; detection;
XX KW identification; cytostatic; cardioactive; neuroprotective; vulnery;
XX KW immunomodulatory; muscular; gynaecological; gastrointestinal;
XX KW nephrotropic; antiinfective; antibacterial; gene therapy; wound;
XX KW neural disorder; immune system disorder; muscular disorder;
XX KW reproductive disorder; gastrointestinal disorder; renal disorder;
XX KW infectious disease; cardiovascular disorder.
XX
```

```
OS Homo sapiens.
XX WO2000055351-A1.
XX 21-SEP-2000.
XX
XX 08-MAR-2000; 2000WO-US05883.
XX PF
XX 12-MAR-1999; 99US-0124270.
XX PR
XX (HUMA-) HUMAN GENOME SCI INC.
XX PA
XX Rosen CA, Ruben SM;
XX PI
XX WPI; 2000-587534/55.
XX DR
XX N-PSDB; AAC98438.
XX
XX Colon cancer associated gene sequences, referred to as colon cancer
XX PT antigens, useful for the treatment, prevention, and diagnosis of colon
XX PT disorders such as colon cancer -
XX PS
XX Claim 11; Page 1810-1811; 2104pp; English.
XX
XX AAC97991 to AAC98763 encode the human colon cancer associated proteins,
XX CC called human colon cancer antigens, given in AAB53234 to AAB54006. The
XX CC human colon cancer antigens can have cytostatic, cardioactive, muscular;
XX CC neuroprotective, immunomodulatory, gynaecological, gastrointestinal,
XX CC vulnery, nephrotropic, antiinfective and antibacterial activities, and
XX CC can be used in gene therapy. The colon cancer antigen polynucleotides,
XX CC proteins and antibodies to the proteins are useful for the prevention,
XX CC treatment and diagnosis of colon disorders, such as colon cancer. The
XX CC polynucleotides may be used in diagnostics and research, such as for
XX CC chromosome identification, and as hybridisation probes. The proteins
XX CC may also be used to prevent diseases such as neural disorders, immune
XX CC system disorders, muscular disorders, reproductive disorders,
XX CC gastrointestinal disorders, wounds, renal disorders, infectious
XX CC diseases, and cardiovascular disorders. AAC98764 to AAC98772 and
XX CC AAB54007 represent sequences used in the exemplification of the present
XX CC invention.
XX
XX Sequence 141 AA;
SQ
Query Match 29.1%; Score 593; DB 21; Length 141;
Best Local Similarity 70.7%; Pred. No. 8.3e-50;
Matches 118; Conservative 7; Mismatches 16; Indels 26; Gaps 2;
Oy 3 DTFIRHIALGFEKRFVPSQHYVYMFVLVKWQDLSEKVVYRRFTEIYEFHKLKEMFPIEA 62
Db 1 dtfirhiallgfekrfvpsqhyv-----hvpgemagpvge 35
Oy 63 GAINPENRIIPHLPAKWFQDQRAAENRGTLTEYCSLTMSLPTKISRCPHLLDFKVP 122
Db 36 ggalpalhrdl-rvpskpwfdgqraaenhqgtlteycgtlmslptkiskrcphlldfkvrp 94
Oy 123 DDLKLPDNTQTKKPTLYLMPKDGKSTATDITGPILQTYRAIANYEK 169
Db 95 ddllkptdnqtkkptelylmpkdgkstatditypiliqtyraianyek 141
RESULT 5
AAG64032
ID AAG64032 standard; protein; 64 AA.
XX
XX AAG64032;
XX
XX 07-SEP-2001 (first entry)
XX
XX Human SH3 domain.
XX
XX Human; antiinflammatory; protein coordinate data; chemical shift data;
XX KW nuclear magnetic resonance; NMR; structural coordinate data;
XX KW PX domain; reduced nicotinamide adenine dinucleotide phosphate; NADPH;
XX
```


XX PA (HYSE-) HYSEQ INC.
XX PI Tang YT, Liu C, Drmanac RT;
XX DR WPI: 2001-451939/48.
XX DR N-PSDB: AAS22479.
XX DR
PT Isolated polypeptides useful for treating anti-inflammatory diseases,
PT nervous system disorders, and for regenerating bone and cartilage -
XX PS Example 4: Page 551-552; 894pp; English.
XX CC The invention relates to polynucleotides encoding novel human
CC proteins or their active domains. The polypeptides, polynucleotides and
CC antibodies raised against the polypeptides are used in a method of
CC treatment of a mammal and prevention of disorders caused by the aberrant
CC protein expression or activity. The polypeptides can be used as
CC molecular weight markers, food supplements, and in antibody production.
CC The polypeptides are used to identify compounds which bind to the
CC polypeptides. Polynucleotides of the invention are used as probes and
CC primers, for sequencing, for chromosome or gene mapping, in the
CC production of recombinant proteins, and in generating anti-sense DNA or
CC RNA and in gene therapy. Polypeptides of the invention can be used to
CC target drugs to a tumour, in assays to determine biological activity, to
CC raise antibodies/ elicit an immune response, to determine quantitative
CC protein levels, as tissue markers, and to isolate receptors or ligands.
CC Polypeptides of the invention may also be useful in treating platelet
CC disorders, stem cell disorders, regenerating bone, cartilage, tendon,
CC ligament and/or nerve tissue, wound healing, treating burns, promoting
CC the proliferation, differentiation and survival of stem cells, as a
CC contraceptive, treating osteoporosis and osteoarthritis, anaemia,
CC Alzheimer's, Parkinson's and Huntington's diseases, amyotrophic lateral
CC sclerosis, stroke, immune deficiencies resulting from bacterial, viral or
CC fungal infection or from autoimmunity, cancer, allergy, asthma,
CC graft-versus-host disease, eczema, haemophilia, thrombosis,
CC anti-inflammatory diseases, nervous system disorders, and infection.
CC The present sequence represents a protein of the invention.
XX PS Sequence 968 AA;
XX QY
Query Match 12.8%; Score 261; DB 22; Length 968;
Best Local Similarity 24.3%; Pred. No. 3.5e-16;
Matches 70; Conservative 46; Mismatches 86; Indels 86; Gaps 7;
QY 156 ILQTYRIANVYKTSSEMAALSTGDDVVEVVEKSESGWFWQAKRGWIPASLEPLDS 215
Db 1 mllqgyvvvnykqnselslqgevvvdielknesgwfvtseegwvpatyleaqqg 60
QY 216 PDETEDEPVPYAG-----EPYVAIKAYTAVEGDEV 245
Db 61 trdsdintstgveksrrkahlrldrrwtlqgmvrqrhsekyvtvqpytsqskdei 120
QY 246 SLLGGEAVEVTHKLLDGW-----KDDVTGYFSPMYLQKSGQDVSAQROI----- 290
Db 121 gfeqgvtvevirknlegwvryrylqke---gwapasyllkakkdlptrkknlagpveilig 177
QY 291 -----KRGAPR-----RSIRNAHSIHQSRKRL 315
Db 178 nimeisnllnkasgdketpaegeheaptakkeislplilcnasngsvagvpdrtvsrl 237
QY 316 SQDAYRNSVRFLOQRQRRARPGQSPGSPLEERQTORSPQPAVPP 363
Db 238 aggs--pavariapraqisnrltrppp-rressigfqlpkppepp 282
RESULT 8
AAW05399
ID AAW05399 standard; Protein: 509 AA.
XX AC AAW05399;
XX

DT 19-FEB-1998 (first entry)
XX Human clone 65 protein.
DE
XX Src-homology region 3 domain; human; mouse; SH3 domain; cell growth;
KW cellular signalling element; cellular structural element; malignancy;
KW protein identification; functional domain; protein screening;
KW cellular signal transduction process.
XX Homo sapiens.
XX OS
XX WO9631625-A1.
XX 10-OCT-1996.
XX PF 04-APR-1996; 96WO-US04454.
XX PR 03-APR-1996; 96US-0630915.
XX PR 07-APR-1995; 95US-0417872.
XX (CYTO-) CYTOGEN CORP.
XX (UYNC-) UNIV NORTH CAROLINA.
XX Fowlkes DM, Hoffman N, Kay BK, McConnell SJ, Sparks AB;
PI WPI: 1996-465045/46.
XX N-PSDB: AAT39799.
XX Identifying polypeptide(s) having specific functional domain (esp.
PT SH3 domain) - comprises detecting selective binding to recognition
PT unit, regardless of sequence homology
XX Claim 54: Fig 59; 174pp; English.
XX AAW05386-W05403 represent novel human and mouse Src-homology region 3
CC (SH3) domain containing proteins that can be used in the method of the
CC invention. SH3 domain containing proteins play a role in signalling and
CC structural elements of cells. The method of the invention is for
CC identifying polypeptides containing functional domains of interest
CC (especially SH3 domains). The method comprises contacting a multivalent
CC recognition unit (RU) complex with a number of peptides and identifying
CC polypeptides having a selective binding affinity for the RU complex. The
CC method is based on functional similarities and does not rely on sequence
CC similarities. Prior methods only gave limited success for identifying
CC proteins which contain an SH3 domain due to the minimal sequence
CC homology among known SH3 proteins. It has been found that small peptide
CC RUS in multivalent form have reduced specificity for a given functional
CC domain compared to monomer RUS. Multivalent RU complexes are particularly
CC suited to screening for polypeptides containing functional domains that
CC are similar to, but not identical in sequence to, the original target
CC functional domain. The new method enables proteins having a common
CC function to be identified. Identification of novel SH3 proteins will be
CC useful for a better understanding of cell growth, malignancy, signal
CC transduction processes, etc. New candidate drugs can be identified, and
CC their specificities (e.g. pharmacological activities) can be assessed
CC using the method of the invention.
XX Sequence 509 AA;
XX QY
Query Match 9.2%; Score 188; DB 17; Length 509;
Best Local Similarity 22.1%; Pred. No. 1.7e-09;
Matches 60; Conservative 58; Mismatches 102; Indels 52; Gaps 9;
QY 64 AINPENRIIPLHPAPKWDGORAENROGTLTEYCSLTMSLPTKISRCPLHLLDFKVRPD 123
Db 141 avspkallp-----ptvslsatstsspslssnqpsvtdyqnvafs 182
QY 124 DLKLPTDNOTKKPETYLMKPGKGTATDITGPI-----ILQTYR--AIANYEKTSGSMA 176
Db 183 nltvntswqkksaft-----rtvpspsvspiHgqgvgvnenlkaqlcswtakdkdnhln 235
QY 177 LSTGDVVEVVEKSESGMWFCQMKAKRGWIPASFLEPLDSPD-ETEDPEPNYA----- 227

Db 236 fskhdilvleqqen-wwfgevhrgrwfkpsvkiilpgsevkrreepealyaavnnkpts 294
Qy 228 -----GEPIYAIKAYTAVGDSVLLGEAEVVIHKLDDGK---DDVTGYFFPSMYLQKS 279
Db 295 aaysvgeeyialypssvepgdlfttegeeilvtqkgewwtgsidrgsifpsnyvkpk 354
Qy 280 QDVVSQAORQIKRGAPRRSRIRNAHSIHORS 311
Db 355 dqesfgsas--ksgasnkkpeiaqvtsayvas 384

RESULT 9
AAB40983
ID AAB40983 standard; Protein: 438 AA.
XX
AC AAB40983;
XX
DT 08-FEB-2001 (first entry)
XX
DE Human ORFX, ORF747 polypeptide sequence SEQ ID NO:1494.
XX
KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW vulnery; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW cardiovascular disease; diabetes mellitus; erythematous; SCID; AIDS;
KW cholesterol ester storage; systemic lupus erythematosus; infection;
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW thrombosis; contraceptive.
XX
OS Homo sapiens.
XX
PN WO200058473-A2.
XX
PD 05-OCT-2000.
XX
PF 31-MAR-2000; 2000WO-US08621.
XX
PR 31-MAR-1999; 99US-0127607.
PR 02-APR-1999; 99US-0127636.
PR 05-APR-1999; 99US-0127728.
PR 30-MAR-2000; 2000US-0540763.
XX
XX (CURA-) CURAGEN CORP.
XX
XX Shinkets RA, Leach M;
PI
XX WPI; 2000-602362/57.
DR N-PSDB; AAC75192.
XX
XX Novel nucleic acids and peptides derived from open reading frame X,
PT useful for treating e.g. cancers, proliferative disorders,
PT neurodegenerative disorders and cardiovascular disease -
XX
XX Claim 11; Page 1240-1241; 5507pp; English.
XX
XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC sequences have activities such as: cytostatic; hepatotropic; vulnery;
CC antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
CC antidiabetic; hypotensive; dermatological; immunosuppressive;
CC antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
CC antithyroid; and antianaemic. The sequences can be used for determining
CC the presence of or predisposition to, or preventing or treating

CC pathological conditions associated with an ORFX-associated disorder. The
CC nucleic acids can be used to express ORFX proteins in gene therapy
CC vectors. The proteins and nucleic acids may be used to treat cancers,
CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
CC graft vs host disease, cardiovascular disease, diabetes mellitus,
CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
CC coagulation; to inhibit thrombosis; and as a contraceptive.
XX
SQ Sequence 438 AA;

Query Match 9.2%; Score 187; DB 21; Length 438;
Best Local Similarity 23.6%; Pred. No. 1.7e-09;
Matches 81; Conservative 47; Mismatches 143; Indels 72; Gaps 15;

Qy 52 KTLKEMFPIEAGAINPENRIIPLHLPAPKWFQDQRAAENROCTLTVEYCSLMSLPTKISRC 111
Db 16 knsgeplppkppgppsh-----pgaldldgvrqnavgrekellssqrd----- 61
Qy 112 PHLLDFFKVRP---DDLK--LPTDNQTKKPEYLMPKDGKSTATDITGPILQIYRAIAN 166
Db 62 ----grfegrpvpgdgakqrskmrqpprrdmtiprglnlppkpppgveeeyytiae 117
Qy 167 YKTSSEMALSTGDVVEVEKSEGWFCOMKAKRGWIPASFLPLELSDPDEDPENY 226
Db 118 fqtptdgisfqaglkveieknlsgwwyiqldkedgwapaftdkykktssnas--fpnf 175
Qy 227 -AGPPY----VAIKAYTAVEGDEVSLLEGEA---VEVIHKLDDG---WKDDVTGYFFPSMY 275
Db 176 laplhvrtqrlgeaaalennttseatgsprlpdpaghyvmdsglpwskdwkg--skdv 233
Qy 276 LQKSGQD-----VSQAORQIKRGAPRRSS--IRNAHSIHORSKRRLSQDAYRRNSV 325
Db 234 lrkassdmsasagyeisdpmeekpslpprkesliksegellerer----- 282
Qy 326 RFLQORRQAR-----PG---PQSPGSPLEERQTKRSKPOP 359
Db 283 ----grteqlrgptkppgvilpmmpakhippardrrrpepkp 321

RESULT 10
AAY71160
ID AAY71160 standard; Protein: 1683 AA.
XX
AC AAY71160;
XX
DT 08-SEP-2000 (first entry)
XX
DE Rat phosphodiesterase interacting protein, M14.
XX
KW Rat; phosphodiesterase interacting protein; M14; PDE; cAMP-PDE;
KW cyclic adenosine monophosphate phosphodiesterase; antiasthmatic; asthma;
KW antiinflammatory; antipsoriatic; dermatological; antibacterial; shock;
KW analgesic; immunosuppressive; antitumor; vasotropic; antiarthritic;
KW antidiabetic; urticaria; antiallergic; antiarteriosclerotic; diagnosis;
KW antirheumatic; treatment; inflammatory disease; psoriasis; arthritis;
KW atherosclerosis; Crohn's disease; cystic fibrosis; chronic bronchitis;
KW eosinophilic granuloma; proliferative skin disease; ulcerative colitis;
KW reperfusion injury; atopic dermatitis; diabetes insipidus;
KW conjunctivitis; adult respiratory distress syndrome; allergic rhinitis;
KW arterial restenosis; ankylosing spondylitis; transplant rejection;
XX graft versus host disease.
XX Rattus sp.
XX WO200027861-A1.
XX
XX 18-MAY-2000.
XX

```
PF 12-NOV-1999; 99WO-US26860.
XX
PR 12-NOV-1998; 98US-0108255.
XX
PA (STRD ) UNIV LELAND STANFORD JUNIOR.
XX
PI Conti M, Pahlke G;
XX
DR WPI; 2000-376479/32.
XX
XX Polynucleotide encoding a phosphodiesterase (PDE) interacting
PT polypeptide, useful for diagnosis and treatment of asthma, cystic
PT fibrosis, Crohn's disease, and rheumatoid arthritis -
XX
PS Disclosure: Fig 6; 77pp; English.
XX
XX The present sequence is a phosphodiesterase (PDE) interacting protein,
CC M14 from rat. The protein modulates the functions and properties of PDEs,
CC specifically cAMP-PDEs, and also targets them to specific subcellular
CC compartments. The present sequence
CC can be used in the diagnosis and treatment of disease conditions
CC associated with PDE activity. The diseases include asthma, cystic
CC fibrosis, inflammatory airway disease, chronic bronchitis, eosinophilic
CC granuloma, psoriasis, proliferative skin diseases, endotoxin shock,
CC septic shock, ulcerative colitis, Crohn's disease, reperfusion injury,
CC inflammatory arthritis, atopic dermatitis, urticaria, adult respiratory
CC distress syndrome, diabetes insipidus, allergic rhinitis, allergic
CC conjunctivitis, vernal conjunctivitis, arterial restenosis,
CC atherosclerosis, inflammatory diseases associated with irritation and
CC pain, rheumatoid arthritis, ankylosing spondylitis, transplant
CC rejection and graft versus host disease, disease conditions associated
CC with hypersecretion of gastric acid, and disease conditions in which
CC cytokines are mediators.
XX
SQ Sequence 1683 AA;

Query Match 9.0%; Score 183; DB 21; Length 1683;
Best Local Similarity 28.0%; Pred. No. 3.2e-08;
Matches 47; Conservative 37; Mismatches 62; Indels 22; Gaps 6;

Qy 162 RAIANEYKTSSEMALSTGDDVVEVSESGWFCQKAKRGWIPASFLPLDSDP-ETE 220
Db qalcswtakkenhlnfskhdvltvedqen--wfgvghgrgwfpxyvkilpgsevkrg 948

Qy 221 DPEPNYA-----GEPYVAIKAYTAVGDEVSLLGEAVEVIHKLDDGWKDDV 267
Db epealyaavnnkktstaypvggeyalysysvpegdltftegeellvtqkdgewtgsi 1008

Qy 268 ---TGYPFSMYLQ-KSGODVSAQRQIKRGAPRRSSIRNAHSIHQRS 311
Db 1009 gertgifsnyvrpkdqnvgnas---ksgasnkkpeiaqvtsayaas 1053

RESULT 11
AAY57445
ID AAY57445 standard; Protein: 1197 AA.
XX
AC AAY57445;
XX
XX 28-FEB-2000 (first entry)
XX
XX Mouse Ese2 protein sequence.
XX
XX Mouse; murine; Ese1; Ese2; endocytosis; vesicular trafficking;
KW regulation; actin cytoskeleton; detection; cancer; infection;
KW LH-domain and SH3-domain regulator of endocytosis; anticancer;
KW antiproliferative; antiviral.
XX
XX Mus sp.
XX
XX WO9955728-A2.
XX
```

```
PD 04-NOV-1999.
XX
XX 27-APR-1999; 99WO-CA00375.
XX
XX 27-APR-1998; 98CA-2230201.
PR 05-FEB-1999; 99US-0118739.
XX
XX (HSCR-) HSC RES & DEV LP.
XX
XX Egan SE, Wang W, Sengar A;
PI
XX WPI; 2000-052802/04.
DR N-PSDB; AAZ39010, AAZ39011.
XX
XX New nucleic acid encoding Ese1 and 2 proteins, involved in regulation
PT of endocytosis, used e.g. for treating cancer or preventing viral
PT infection -
XX
XX Claim 33; Page 48; 99pp; English.
XX
```

```
CC The present sequence represents mouse Ese2. The present invention
CC specifically describes mammalian Ese1 and 2 proteins (I) and their splice
CC variants (Ese - EH-domain and SH3-domain regulator of endocytosis). (I)
CC are involved in regulation of clathrin-mediated endocytosis (as a complex
CC with Esp15 protein), vesicular trafficking and actin cytoskeleton.
CC Generally (I) or its (ant)agonists, mimetics, fragments and inactive
CC mutants; (I)-specific antibodies (Ab); sequences antisense to the (I)
CC polynucleotide; agents that downregulate expression of Ese genes or
CC antagonists of an Ese binding partner are used to treat diseases
CC associated with undesirable endocytosis and resulting changes in cellular
CC function. Particularly overexpression of Ese1 is used to block
CC clathrin-mediated endocytosis in vivo or in cell cultures, while
CC administration of (I) is used to promote endocytosis of selected cells.
CC (Ant)agonists of (I) or Ab are used to suppress abnormal proliferation of
CC cells that can be stimulated to proliferate by a growth factor receptor;
CC and similar compounds (also inactive Ese mutants) can be used to prevent
CC viral infection. Endocytosis may also be regulated, in vivo or in cell
CC cultures, by forming an Ese-Esp15 complex, then binding dynamitin to the
CC complex. Generally conditions that can be treated include cancer;
CC abnormal cell division or migration; viral infection; or abnormal
CC receptor signalling, tissue development or synaptic transmission.
XX
SQ Sequence 1197 AA;
```

```
Query Match 8.7%; Score 177.5; DB 21; Length 1197;
Best Local Similarity 23.2%; Pred. No. 6.6e-08;
Matches 63; Conservative 50; Mismatches 108; Indels 51; Gaps 10;
```

```
Qy 80 WF-----DGORAAENRQGLTEYCSLMSLPTKISRCP-HLLDFKVRPDDKLPT 129
Db 767 wfpcnyvekvlssekalspkallpptvs--lsatstssqppasvtdyhvnvfnslvtnt 824

Qy 130 DNQTKKPET-----YLMPPDKGSTATDITGPIILQYRAIANEYKTSSEMALSTGDDV 183
Db 825 twgkksaftvtvsgvsvphggggavenl-----kaqalcswtakkenhlnfskhdvi 878

Qy 184 EVVEKSESGWFCQKAKRGWIPASFLPLDSDP-ETEDPEPNYA----- 227
Db 879 tvleqqen--wfgvghgrgwfpxyvkilpgnevgrgepealyaavtckktpstapvts 937

Qy 228 -----GEPYVAIKAYTAVGDEVSLLGEAVEVIHKLDDGWKDDV---TGYPFSMYLQKS 279
Db 938 taypvgedyalysysvpegdltftegeellvtqkdgewtgsigertgifsnyvrpk 997

Qy 280 QGDVSAQRQIKRGAPRRSSIRNAHSIHQRS 311
Db 998 dge--nfgnasksgasnkkpeiaqvtsayaas 1027
```

```
RESULT 12
AAY57450
ID AAY57450 standard; Protein: 1658 AA.
```


XX
PN
W09955728-A2

PN WO995

XX PN WO9955728-A2.

```
XX 04-NOV-1999.
PD
XX 27-APR-1999; 99WO-CA00375.
PF
XX 27-APR-1998; 98CA-2230201.
PR
XX 05-FEB-1999; 99US-0118739.
PR
XX (HSCR-) HSC RES & DEV LP.
PA
XX Egan SE, Wang W, Sengar A;
PI
XX WPI; 2000-052802/04.
XX N-PSDB; AAZ39024, AAZ39025.
DR
XX New nucleic acid encoding Esel and 2 proteins, involved in regulation
PT of endocytosis, used e.g. for treating cancer or preventing viral
PT infection
PT
XX Claim 14; Page 62-63; 99pp; English.
PS
XX The present invention specifically describes mammalian Esel and 2
CC proteins (I) and their splice variants (Ese = EH-domain and SH3-domain
CC regulator of endocytosis). (I) are involved in regulation of clathrin-
CC mediated endocytosis (as a complex with Esp15 protein), vesicular
CC trafficking and actin cytoskeleton. Generally (I) (or its (ant)agonists,
CC mimetics, fragments and inactive mutants); (I)-specific antibodies (Ab);
CC sequences antisense to the (I) polynucleotide; agents that downregulate
CC expression of Ese genes or antagonists of an Ese binding partner are
CC used to treat diseases associated with undesirable endocytosis and
CC resulting changes in cellular function. Particularly overexpression of
CC Ese1 is used to block clathrin-mediated endocytosis in vivo or in cell
CC cultures, while administration of (I) is used to promote endocytosis of
CC selected cells. (Ant)agonists of (I) or Ab are used to suppress abnormal
CC proliferation of cells that can be stimulated to proliferate by a growth
CC factor receptor; and similar compounds (also inactive Ese mutants) can be
CC used to prevent viral infection. Endocytosis may also be regulated, in
CC vivo or in cell cultures, by forming an Ese-Esp15 complex, then binding
CC dynamin to the complex. Generally conditions that can be treated include
CC cancer; abnormal cell division or migration; viral infection; or abnormal
CC receptor signalling, tissue development or synaptic transmission. The
CC present sequence represents mouse Ese1L protein sequence.
XX
XX Sequence 1715 AA;
SQ
Query Match 8.3%; Score 168.5; DB 21; Length 1715;
Best Local Similarity 31.2%; Pred. No. 8.6e-07;
Matches 54; Conservative 34; Mismatches 58; Indels 27; Gaps 11;
QY 136 PETYLMRPDGRKSTAP-DITG-----PILOTYRAIANYEKTSGSEMALSTGDVVEVVEKS 189
Db 1045 pnyvrlkdsgsgtagtsglgkkpeiaq---viasyaatgpeqiltapqqlilirkkn 1101
QY 190 RSGWFWCOMKA---KR--GWIPASGLEPLDSPDE---TEDPEPNYAGEP----YVAIKAY 237
Db 1102 pggwweqelqargkkrqigwfpnyvkil-spqtskitptelptavqpvccvqigmydy 1160
QY 238 TAVEGDEVSLLEGAVEVIHKL-LDGWKKDDVT---GYFSPMYLQ-KSCQDVYQ 285
Db 1161 taqnddelafskqqlinvinkepdwkwgevsqvgvlfpsnyvkittmdpsq 1213
```

Search completed: January 11, 2002, 09:08:48
Job time: 174 sec